Bisimulation and expressivity for conditional belief, degrees of belief, and safe belief

Andersen, Mikkel Birkegaard; Bolander, Thomas; van Ditmarsch, Hans; Jensen, Martin Holm

Published in:
Synthese

Link to article, DOI:
10.1007/s11229-016-1060-x

Publication date:
2016

Document Version
Peer reviewed version

Bisimulation and expressivity for conditional belief, degrees of belief, and safe belief

Mikkel Birkegaard Andersen[*], Thomas Bolander,
Hans van Ditmarsch[†], and Martin Holm Jensen

February 29, 2016

Abstract

Plausibility models are Kripke models that agents use to reason about knowledge and belief, both of themselves and of each other. Such models are used to interpret the notions of conditional belief, degrees of belief, and safe belief. The logic of conditional belief contains that modality and also the knowledge modality, and similarly for the logic of degrees of belief and the logic of safe belief. With respect to these logics, plausibility models may contain too much information. A proper notion of bisimulation is required that characterises them. We define that notion of bisimulation and prove the required characterisations: on the class of image-finite and preimage-finite models (with respect to the plausibility relation), two pointed Kripke models are modally equivalent in either of the three logics, if and only if they are bisimilar. As a result, the information content of such a model can be similarly expressed in the logic of conditional belief, or the logic of degrees of belief, or that of safe belief. This, we found a surprising result. Still, that does not mean that the logics are equally expressive: the logics of conditional and degrees of belief are incomparable, the logics of degrees of belief and safe belief are incomparable, while the logic of safe belief is more expressive than the logic of conditional belief. In view of the result on bisimulation characterisation, this is an equally surprising result. We hope our insights may contribute to the growing community of formal epistemology and on the relation between qualitative and quantitative modelling.

1 Introduction

A typical approach in belief revision involves preferential orders to express degrees of belief and knowledge [20, 27]. This goes back to the 'systems of spheres' in [25, 16]. Dynamic doxastic logic was proposed and investigated in [28] in order to provide a link between the (non-modal logical) belief revision and modal logics with explicit knowledge and belief operators. A similar approach was pursued in belief revision in dynamic epistemic logic [5, 36, 32, 8, 39], that continues to develop strongly [12, 33]. We focus on the proper notion of structural equivalence on models encoding knowledge and belief simultaneously. A prior investigation into that is [13], which we relate our results to at the end of the paper. Our motivation is to find suitable structural notions to reduce the complexity of solving planning problems. Solutions to planning problems are sequences of actions, such as iterated belief revision. It is the dynamics of knowledge and belief that, after all, motivates our research.

[*]DTU Compute, Technical University of Denmark, {mibi,tobo,mhje}@dtu.dk
[†]LORIA, CNRS / Université de Lorraine, hans.van-ditmarsch@loria.fr

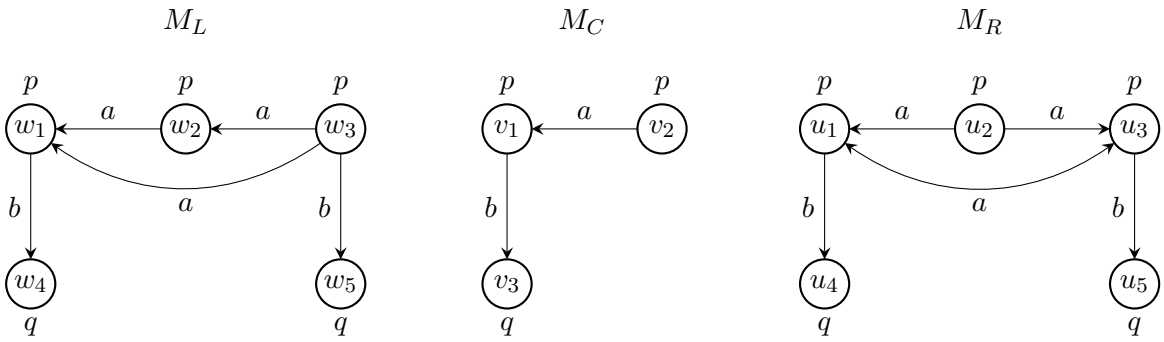

Figure 1: An arrow $x \to y$ labelled by a means $x \geq_a y$; agent a considers y at least as plausible as x. We use $x >_a y$ to mean $x \geq_a y$ and $y \not\geq_a x$. Here $w_2 >_a w_1$, so w_1 is strictly more plausible than w_2. Reflexive edges are omitted. Unlisted propositions are false.

The semantics of belief depends on the structural properties of models. To relate the structural properties of models to a logical language we need a notion of structural similarity, known as bisimulation. A bisimulation relation relates a modal operator to an accessibility relation. Plausibility models do not have an accessibility relation as such but a plausibility relation. This induces a set of accessibility relations: the *most plausible* worlds are the *accessible* worlds for the modal belief operator; and the *plausible* worlds are the *accessible* worlds for the modal knowledge operator. But it contains much more information: to each modal operator of conditional belief (or of degree of belief) one can associate a possibly distinct accessibility relation.This raises the question of how to represent the bisimulation conditions succinctly. Can this be done by reference to the plausibility relation directly, instead of by reference to these, possibly many, induced accessibility relations? It is now rather interesting to observe that relative to the modalities of knowledge and belief, the plausibility relation is already in some way too rich.

The plausibility model M_L on the left in Figure 1 consists of five worlds. The proposition p is true in the top ones and false in the bottom ones. The reverse holds for q: true at the bottom and false at the top. The a relations in the model correspond to the plausibility order $w_3 >_a w_2 >_a w_1$, interpreted such that the smaller of two elements in the order is the most plausible of the two. Further, everything that is comparable with the plausibility order is considered epistemically possible. Hence, the epistemic equivalence classes for agent a in M_L are $\{w_1, w_2, w_3\}$, $\{w_4\}$ and $\{w_5\}$. We can then view the model as a standard multi-agent $S5$ model plus an ordering on the epistemic possibilities. As w_1 is the most plausible world for a in the equivalence class $\{w_1, w_2, w_3\}$, she will in w_3 believe p and that b believes $\neg p \wedge q$. This works differently from the usual doxastic modal logic, where belief corresponds to the accessibility relation. In the logics of belief that we study, belief is what holds in the most plausible world(s) in an epistemic equivalence class. For a, the most plausible world in her equivalence class $\{w_1, w_2, w_3\}$ is w_1, so a believes the same formulas in all of them.

In w_2 agent b knows p. If a is given the information that b does not consider q possible (that is, the information that neither w_1 nor w_3 is the actual world), then a believes that b knows p – or conditional on $K_b\neg q$, a believes $K_b p$. Such a statement is an example of the logic of conditional belief L^C defined in Section 3. In L^C we write this statement as $B_a^{K_b\neg q} K_b p$.

Now examine w_3. We will show that w_1 and w_3 are modally equivalent for L^C: they agree on all formulas of that language—no information expressible in L^C distinguishes the

two worlds. This leads to the observation that no matter where we move w_3 in the plausibility ordering for a, modal equivalence is preserved. Similarly, we can move w_2 anywhere we like *except* making it more plausible than w_1. If we did, then a would believe $K_b p$ unconditionally, and the formulas true in the model would have been changed.

It turns out that moving worlds about in the plausibility order can be done for all models, as long as we obey one (conceptually) simple rule: Grouping worlds into "modal equivalence classes" of worlds modally equivalent to each other, we are only required to preserve the ordering between the *most* plausible worlds in each modal equivalence class. *Only the most plausible world in each class matters.*

Another crucial observation is that standard bisimulation in terms of $\geq_a$ does not give correspondence between bisimulation and modal equivalence. For instance, while w_1 and w_3 are modally equivalent, they are not "standardly" bisimilar with respect to $\geq_a$: w_3 has a $\geq_a$-edge to a $K_b p$ world (w_2), whereas w_1 does not. Thus, the straightforward, standard definition of bisimulation does not work, because no modality in the language corresponds to the plausibility relation. Instead we have an infinite set of modalities corresponding to relations derived from the plausibility relation. One of the major contributions of this paper is a solution to exactly this problem.

Making w_3 as plausible as w_1 and appropriately renaming worlds gets us M_R of Figure 1. Here the modally equivalent worlds u_1 and u_3 are equally plausible, modally equivalent *and* standardly bisimilar. This third observation gives a sense of how we solve the problem generally. Rather than using $\geq_a$ directly, our definition of bisimulation checks accessibility with respect to a relation $\geq_a^R$ derived from $\geq_a$ and the bisimulation relation R itself. Postponing details for later we just note that in the present example the derived relation for M_L is exactly the plausibility relation for M_R. This indicates what we later prove: This new derived relation reestablishes the correspondence between bisimilarity and modal equivalence.

The model M_C of Figure 1 is the bisimulation contraction of the right model using standard bisimilarity. It is the bisimulation contraction of both models with the bisimulation notion informally defined in the previous paragraph. In previous work on planning with single-agent plausibility models [2], finding contractions of plausibility models is needed for decidability and complexity results. In this paper we do this for the first time for multi-agent plausibility models, opening new vistas in applications of modal logic to automated planning.

Overview of content In Section 2 we introduce plausibility models and the proper and novel notion of bisimulation on these models, and prove various properties of bisimulation. In Section 3 we define the three logics of conditional belief, degrees of belief, and safe belief, and provide some further historical background on these logics. In Section 4 we demonstrate that bisimilarity corresponds to logical equivalence (on image-finite and preimage-finite models) for all three core logics, so that, somewhat remarkably, one could say that the content of a given model can equally well be described in any of these logics. Then, in Section 5 we determine the relative expressivity of the three logics, including more expressive combinations of their primitive modalities. The main result here is that the logics of conditional and degrees of belief are incomparable, and that the logics of degrees of belief and safe belief are incomparable, but that the logic of safe belief is (strictly) more expressive than the logic of conditional belief. In Section 6, we put our result in the perspective of other recent investigations, mainly the study by Lorenz Demey [13], and in the perspective of possible applications: decidable planning.

2 Plausibility models and bisimulation

A *well-preorder* on a set X is a reflexive and transitive binary relation $\trianglerighteq$ on X such that every non-empty subset has $\trianglerighteq$-minimal elements. The set of *minimal elements* (for $\trianglerighteq$) of some $Y \subseteq X$ is the set $Min_{\trianglerighteq} Y$ defined as $\{y \in Y \mid y' \trianglerighteq y \text{ for all } y' \in Y\}$.[1] As any two-element subset $Y = \{x, y\}$ of X also has minimal elements, we have that $x \trianglerighteq y$ or $y \trianglerighteq x$. Thus all elements in X are $\trianglerighteq$-comparable.

Given any binary relation R on X, we use $R^{=}$ to denote the reflexive, symmetric, and transitive closure of R (the equivalence closure of R). For any equivalence relation R on X, we write $[x]_R$ for $\{x' \in X \mid (x, x') \in R\}$. A binary relation R on X is *image-finite* if and only if for every $x \in X$, $\{x' \in X \mid (x, x') \in R\}$ is finite. A relation is *preimage-finite* if and only if for every $x \in X$, $\{x' \in X \mid (x', x) \in R\}$ is finite. We say R is *(pre)image-finite* if it is both image-finite and preimage-finite. We often write xRy instead of $(x, y) \in R$. Given subsets $Y, Z \subseteq X$, we define YRZ if and only if yRz for all $y \in Y$ and all $z \in Z$.

Definition 1 (Plausibility model)**.** A *plausibility model* for a countably infinite set of propositional symbols P and a finite set of agents A is a tuple $M = (W, \geq, V)$, where

- W is a set of *worlds* called the *domain*, denoted $D(M)$;
- $\geq$: $A \to \mathcal{P}(W \times W)$ is a function mapping each $a \in A$ into a *plausibility relation* $\geq(a)$, usually abbreviated $\geq_a$. For each $a \in A$ and $w \in W$, $\geq_a$ is a well-preorder on the set $\{w' \in W \mid w \geq_a w' \text{ or } w' \geq_a w\}$. Each $\geq_a$ is required to be (pre)image-finite;
- $V : W \to 2^P$ is a *valuation*.

For $w \in W$, (M, w) is a *pointed plausibility model*.

If $w \geq_a v$ then v is *at least as plausible* as w (for agent a), and the $\geq_a$-minimal elements are the *most plausible* worlds. For the symmetric closure of $\geq_a$ we write $\sim_a$: this is an equivalence relation on W called the *epistemic relation* (for agent a). If $w \geq_a v$ but $v \not\geq_a w$ we write $w >_a v$ (v is *more plausible* than w), and for $w \geq_a v$ and $v \geq_a w$ we write $w \simeq_a v$ (w and v are *equiplausible*). Instead of $w \geq_a v$ ($w >_a v$) we may write $v \leq_a w$ ($v <_a w$).

Note that we have required each relation $\geq_a$ to be (pre)image-finite. This amounts to requiring that all equivalence classes of $\sim_a$ are finite, while still allowing infinite domains. This requirement is not part of the definition of plausibility models provided in [8]. We require it here, since it leads to simplifications without any significant reduction in generality:

1. We will show full correspondence between bisimilarity and modal equivalence for three different logics over plausibility models. As is the case in standard modal logic, this correspondence can only be achieved for (pre)image-finite models (the direction from modal equivalence to bisimilarity only hold for such models, see e.g. [9]). Simply assuming (pre)image-finiteness from the outset simplifies the presentation, as we do not have to repeat this restriction in a large number of places.

2. Some of our later results are going to rely on the existence of a largest autobisimulation (see below). Usually it is quite trivial to show the existence of a largest bisimulation,

[1] This notion of minimality is non-standard and taken from [8]. Usually a minimal element of a set is an element that is not greater than any other element.

since the union of any set of bisimulations is a bisimulation. However, we need a non-standard notion of bisimulation for our purposes, and for such bisimulations closure under union is far from a trivial result.[2] Given our first correspondence result between bisimilarity and modal equivalence, we are however going to get this result for free (see Section 4.1).

We now proceed to define a notion of autobisimulation on a plausibility model. This notion is non-standard, because there is no one-to-one relation between the plausibility relation for an agent and a modality for that agent in the logics defined later. In the definition below (and from now on), we allow ourselves some further notational abbreviations. Let $M = (W, \geq, V)$ denote a plausibility model. Let $a \in A$ and $w \in W$, then we write $[w]_a$ instead of $[w]_{\sim_a}$. Now let $Z \subseteq [w]_a$, then we write $Min_a\, Z$ instead of $Min_{\geq_a} Z$. For any binary relation R on W, we write $w \geq_a^R v$ for $Min_a([w]_{R^=} \cap [w]_a) \geq_a Min_a([v]_{R^=} \cap [v]_a)$. When $w \geq_a^R v$ and $v \geq_a^R w$, we write $w \simeq_a^R v$.

Definition 2 (Autobisimulation)**.** Let $M = (W, \geq, V)$ be a plausibility model. An *autobisimulation* on M is a non-empty relation $R \subseteq W \times W$ such that for all $(w, w') \in R$ and for all $a \in A$:

[atoms] $V(w) = V(w')$;

[forth$_\geq$] If $v \in W$ and $w \geq_a^R v$, there is a $v' \in W$ such that $w' \geq_a^R v'$ and $(v, v') \in R$;

[back$_\geq$] If $v' \in W$ and $w' \geq_a^R v'$, there is a $v \in W$ such that $w \geq_a^R v$ and $(v, v') \in R$;

[forth$_\leq$] If $v \in W$ and $w \leq_a^R v$, there is a $v' \in W$ such that $w' \leq_a^R v'$ and $(v, v') \in R$;

[back$_\leq$] If $v' \in W$ and $w' \leq_a^R v'$, there is a $v \in W$ such that $w \leq_a^R v$ and $(v, v') \in R$.

A *total autobisimulation* on M is an autobisimulation with W as both domain and codomain.

Our bisimulation relation is non-standard in the [back] and [forth] clauses. A standard [forth] condition based on an accessibility relation $\geq_a$ would be

If $v \in W$ and $w \geq_a v$, there is a $v' \in W'$ such that $w' \geq_a v'$ and $(v, v') \in R$.

Here, R only appears in the part '$(v, v') \in R$'. But in the definition of autobisimulation for plausibility models, in [forth$_\geq$], the relation R also features in the condition for applying [forth$_\geq$] and in its consequent, namely as the upper index in $w \geq_a^R v$ and $w' \geq_a^R v'$. This means that R also determines which v are accessible from w, and which v' are accessible from w'. This explains why we define an autobisimulation on a single model before a bisimulation between distinct models: We need the bisimulation relation R to determine the plausibility relation $\geq_a^R$ from the plausibility relation $\geq_a$ on any given model first, before structurally comparing distinct models.

Example 1. The models M_L and M_R of Figure 1 are reproduced in Figure 2. Consider the relation $R = R_{id} \cup \{(w_1, w_3), (w_3, w_1), (w_4, w_5), (w_5, w_4)\}$, where R_{id} is the identity relation

[2]Without the restriction to (pre)image-finite models, we were unable to prove the existence of a largest bisimulation. We leave this challenge open to future research(ers).

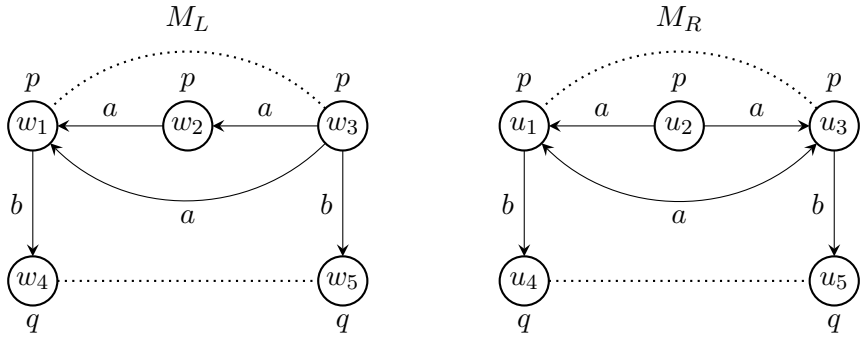

Figure 2: The left and right models of Figure 1, with the dotted lines showing the largest autobisimulations (modulo reflexivity).

on W. With this R, we get that w_1 and w_3 are equiplausible for $\geq_a^R$:

$$\begin{aligned} w_1 \simeq_a^R w_3 \text{ iff} \\ Min_a([w_1]_{R^=} \cap [w_1]_a) \simeq_a Min_a([w_3]_{R^=} \cap [w_3]_a) \text{ iff} \\ Min_a\{w_1, w_3\} \simeq_a Min_a\{w_1, w_3\} \text{ iff} \\ w_1 \simeq_a w_1 \end{aligned}$$

We also get that $w_2 \geq_a^R w_3$:

$$\begin{aligned} w_2 \geq_a^R w_3 \text{ iff} \\ Min_a([w_2]_{R^=} \cap [w_2]_a) \geq_a Min_a([w_3]_{R^=} \cap [w_3]_a) \text{ iff} \\ Min_a\{w_2\} \geq_a Min_a\{w_1, w_3\} \text{ iff} \\ w_2 \geq_a w_1 \end{aligned}$$

This gives $\geq_a^R = \{(w_1, w_3), (w_3, w_1), (w_2, w_3), (w_2, w_1)\} \cup R_{id}$. For b, we get $\geq_b^R = \geq_b$. The autobisimulation R on M_L is shown in Figure 2. It should be easy to check that R is indeed an autobisimulation. To help, we will justify why (w_4, w_5) is in R: For $\geq_b^R$, we have that, as $(w_1, w_3) \in R$ and $w_1 \geq_b^R w_4$, there must be a world v such that $w_3 \geq_b^R v$ and $(w_4, v) \in R$. This v is w_5.

Note that R is the largest autobisimulation. Based on [atoms] there are only two possible candidate pairs that could potentially be added to R (modulo symmetry), namely (w_1, w_2) and (w_2, w_3). But w_2 does not have a b-edge to a q world, whereas both w_1 and w_3 do. There is therefore nothing more to add.

The largest autobisimulation for M_R is completely analogous, as shown in Figure 2.

Lemma 1. *Let $M = (W, \geq, V)$ be a plausibility model and R a binary relation on W. If $(w, w') \in R^=$ and $w \sim_a w'$ then $w \simeq_a^R w'$.*

Proof. From $(w, w') \in R^=$ and $w \sim_a w'$ we get $[w]_{R^=} = [w']_{R^=}$ and $[w]_a = [w']_a$ and hence $[w]_{R^=} \cap [w]_a = [w']_{R^=} \cap [w']_a$. Thus also $Min_a([w]_{R^=} \cap [w]_a) = Min_a([w']_{R^=} \cap [w']_a)$, immediately implying $w \simeq_a^R w'$. □

Let $M = (W, \geq, V)$ be a plausibility model. By definition, for each agent a, $\geq_a$ is a well-preorder on each $\sim_a$-equivalence class. The following result shows that the same holds for $\geq_a^R$ where R is *any* binary relation.

Lemma 2. *Let $M = (W, \geq, V)$ be a plausibility model and R a binary relation on M. Then $\geq_a^R$ is a well-preorder on each $\sim_a$-equivalence class.*

Proof. The relation $\geq_a$ partitions W into a well-preorder on each $\sim_a$-equivalence class, by definition. We need to show that $\geq_a^R$ does the same. Hence we need to prove: 1) $\geq_a^R$ is reflexive; 2) $\geq_a^R$ is transitive; 3) any $\sim_a$-equivalence class has $\geq_a^R$-minimal elements; 4) if two worlds are related by $\geq_a^R$ they are also related by $\sim_a$.

Reflexivity of $\geq_a^R$ is trivial. *Transitivity*: Let $(w, v), (v, u) \in \geq_a^R$. Then $Min_a([w]_{R^=} \cap [w]_a) \geq_a Min_a([v]_{R^=} \cap [v]_a)$, and $Min_a([v]_{R^=} \cap [v]_a) \geq_a Min_a([u]_{R^=} \cap [u]_a)$. Using that for any sets X, Y, Z, if $X \geq_a Y$ and $Y \geq_a Z$ then $X \geq_a Z$ (transitivity of $\geq_a$ for sets is easy to check), we obtain that $Min_a([w]_{R^=} \cap [w]_a) \geq_a Min_a([u]_{R^=} \cap [u]_a)$ and therefore $(w, u) \in \geq_a^R$. *Minimal elements*: Consider a $\sim_a$-equivalence class $W'' \subseteq W$, and let $W' \subseteq W''$ be a non-empty subset. Suppose W' does not have $\geq_a^R$ minimal elements. Then for all $w' \in W'$ there is a $w'' \in W'$ such that $w'' <_a^R w'$, i.e. $Min_a([w'']_{R^=} \cap [w'']_a) <_a Min_a([w']_{R^=} \cap [w']_a)$. As $w' \in [w']_{R^=} \cap [w']_a$, we get $\{w'\} \geq_a Min_a([w']_{R^=} \cap [w']_a)$ and then $Min_a([w'']_{R^=} \cap [w'']_a) <_a \{w'\}$. In other words, for all $w' \in W'$ there is a $u \in W$, namely any $u \in Min_a([w'']_{R^=} \cap [w'']_a)$, such that $u <_a w'$. This contradicts $\geq_a$ being a well-preorder on W''. We have now shown 1), 2) and 3). Finally we show 4): Assume $w \geq_a^R v$, that is, $Min_a([w]_{R^=} \cap [w]_a) \geq_a Min_a([v]_{R^=} \cap [v]_a)$. This implies the existence of an $x \in Min_a([w]_{R^=} \cap [w]_a)$ and a $y \in Min_a([w]_{R^=} \cap [v]_a)$ with $x \geq_a y$. By choice of x and y we have $x \sim_a w$ and $y \sim_a v$. From $x \geq_a y$ we get $x \sim_a y$. Hence we have $w \sim_a x \sim_a y \sim_a v$, as required. □

Proposition 1. *On any plausibility model there exists a largest autobisimulation. Furthermore, the largest autobisimulation is an equivalence relation.*

We postpone the proof of this result to Section 4.1, since it is going to follow from the correspondence between bisimilarity and modal equivalence for our language of conditional belief (Theorem 1). Let us already now reassure the reader that we are not risking any circular reasoning here: None of the results that lead to Theorem 1 and hence to Proposition 1 rely on largest autobisimulations.

Definition 3 (Bisimulation)**.** Let $M = (W, \geq, V)$ and $M' = (W', \geq', V')$ be plausibility models and let $M'' = M \sqcup M'$ be the disjoint union of the two. Given an autobisimulation R on M'', if $R' = R \cap (W \times W')$ is non-empty, then R' is called a *bisimulation* between M and M'. A bisimulation between (M, w) and (M', w') is a bisimulation between M and M' containing (w, w').

Example 2. Take another look at M_C and M_R of Figure 1. Let $M' = M_C \sqcup M_R$ and consider possible autobisimulations here. From Figure 2 we have the existence of a largest autobisimulation on M_R. For M_C, the largest autobisimulation is just the identity. Naming them R_R and R_C respectively, we (trivially) have that $R_R \cup R_C$ is an autobisimulation on M'. The question is whether we can extend $R_R \cup R_C$ to an autobisimulation on M' connecting the submodels M_R to M_C. We can. This new autobisimulation is $R = R' \cup R_R \cup R_C$, where $R'(u_1) = R'(u_3) = \{v_1\}$, $R'(u_2) = \{v_2\}$ and $R'(u_4) = R'(u_5) = \{v_3\}$. Now we easily get a bisimulation between M_R and M_C as $R \cap (D(M_R) \times D(M_C)) = R'$. Figure 3 shows the bisimulation R'.

Definition 4 (Bisimulation contraction)**.** Let $M = (W, \geq, V)$ be a plausibility model and let R be the largest autobisimulation on M. The *bisimulation contraction* of M is the model

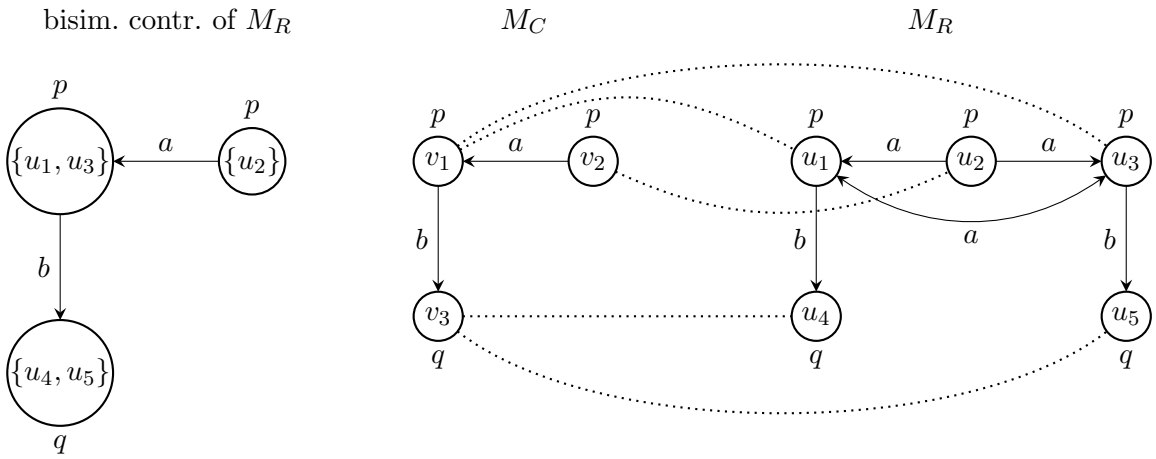

Figure 3: See Figure 1. The dotted edges show the largest bisimulation between M_C and M_R. Model M_C is isomorphic to the bisimulation contraction of M_R (on the left) and to the bisimulation contraction of M_L (not depicted).

$M' = (W', \geq', V')$ such that $W' = \{[w]_R \mid w \in W\}$, $V'([w]_R) = V(w)$, and for all agents a and worlds $w, v \in W$:

$$[w]_R \geq'_a [v]_R \quad \text{iff} \quad \text{for some } w' \in [w]_R \text{ and } v' \in [v]_R\text{: } w' \geq^R_a v'.$$

Example 3. We compute the simulation contraction $M'_R = (W', \geq', V')$ of $M_R = (W, \geq, V)$. For $\geq'_a$ and $\geq'_b$ take the reflexive closures.

$$\begin{array}{lcl} W' & = & \{\{u_1, u_3\}, \{u_2\}, \{u_4, u_5\}\} \\ \geq'_a & = & \{(\{u_2\}, \{u_1, u_3\})\} \\ \geq'_b & = & \{\{u_1, u_3\}, \{u_4, u_5\})\} \\ V'(\{u_1, u_3\}) & = & \{p\} \\ V'(\{u_2\}) & = & \{p\} \\ V'(\{u_4, u_5\}) & = & \{q\} \end{array}$$

Model M_C is isomorphic to both the bisimulation contraction of M_L and the bisimulation contraction of M_R.

Proposition 2. *The bisimulation contraction of a plausibility model is a plausibility model and is bisimilar to that model.*

This proposition is not hard to prove, so we do not provide a full proof, but only sketch the overall idea. First, to prove that the bisimulation contraction $(W', \geq', V')$ of a plausibility model $(W, \geq, V)$ is a plausibility model, we simply have to prove that the relations $\geq'_a$ are well-preorders on each $\sim'_a$ equivalence class. That is shown by first proving reflexivity of $\geq'_a$, then transitivity of $\geq'_a$, and finally by proving that any non-empty subset has minimal elements with respect to $\geq'_a$. To show that $(W, \geq, V)$ is bisimilar to $(W', \geq', V')$, we define the (functional) relation $S : W \to W'$ as $S = \{(w, [w]_R) \mid w \in W\}$ and show that this is a bisimulation relation (that it satisfied [atoms] and the [back] and [forth] conditions).

Definition 5 (Normal plausibility relation, normal model)**.** Let $M = (W, \geq, V)$ be a plausibility model and let R be the largest autobisimulation on M. For all agents a, the relation

$\geq_a^R$ is the *normal plausibility relation* for agent a in M, for which we may also write $\succeq_a$. The model is *normal* if for all a, $\geq_a = \geq_a^R$. Any model M can be *normalised* by replacing all $\geq_a$ by $\geq_a^R$.

Example 4. Consider again the models M_L and M_R of Figure 2. From the largest bisimulation on M_L (shown by dotted edges), we can conclude that M_R is the normalisation of M_L (modulo a renaming of the worlds w_i to u_i, for $i = 1, \ldots, 5$).

Proposition 3. *The bisimulation contraction of a plausibility model is normal.*

Proof. Let M be a plausibility model, and let $M' = (W', \geq', V')$ be the bisimulation contraction of M. The largest autobisimulation on M' is the identity relation R_{id}. For each agent a, we now have that $\geq_a'^{R_{id}} = \geq_a'$. Therefore, M' is normal. □

3 Logical language and semantics

In this section we define the language and the semantics of our logics.

Definition 6 (Logical language)**.** For any countably infinite set of propositional symbols P and finite set of agents A we define language L_{PA}^{CDS} by:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_a\varphi \mid B_a^{\varphi}\varphi \mid B_a^n\varphi \mid \Box_a\varphi$$

where $p \in P$, $a \in A$, and $n \in \mathbb{N}$.

The formula $K_a\varphi$ stands for 'agent a knows (formula) φ', $B_a^{\psi}\varphi$ stands for 'agent a believes φ on condition ψ', $B_a^n\varphi$ stands for 'agent a believes φ to degree n', and $\Box_a\varphi$ stands for 'agent a safely believes φ'. (The semantics of these constructs is defined below.) The duals of K_a, B_a^{φ} and $\Box_a$ are denoted $\widehat{K}_a$, $\widehat{B}_a^{\varphi}$ and $\Diamond_a$. We use the usual abbreviations for the boolean connectives as well as for $\top$ and $\bot$, and the abbreviation B_a for $B_a^{\top}$. In order to refer to the type of modalities in the text, we call K_a a *knowledge modality*, B_a^{ψ} a *conditional belief modality*, B_a^n a *degree of belief modality*, and $\Box_a$ a *safe belief modality*.

In L_{PA}^{CDS}, if A is clear from the context, we may omit that and write L_P^{CDS}, and if P is clear from the context, we may omit that as well, so that we get L^{CDS}. The letter C stands for 'conditional', D for 'degree', and S for 'safe'. Let X be any subsequence of CDS, then L^X is the language with, in the inductive definition, only the modalities X (and with knowledge K_a) for all agents. In our work we focus on the *logic of conditional belief* with language L^C, the *logic of degrees of belief* with language L^D, and the *logic of safe belief* with language L^S.

Definition 7 (Satisfaction Relation)**.** Let $M = (W, \geq, V)$ be a plausibility model for P and A, let $\succeq$ be the normal plausibility relation for M, and let $w \in W$, $p \in P$, $a \in A$, and $\varphi, \psi \in L^{CDS}$. Then:

$$\begin{array}{lll}
M, w \models p & \text{iff} & p \in V(w) \\
M, w \models \neg\varphi & \text{iff} & M, w \not\models \varphi \\
M, w \models \varphi \wedge \psi & \text{iff} & M, w \models \varphi \text{ and } M, w \models \psi \\
M, w \models K_a\varphi & \text{iff} & M, v \models \varphi \text{ for all } v \in [w]_a \\
M, w \models B_a^{\psi}\varphi & \text{iff} & M, v \models \varphi \text{ for all } v \in Min_a([\![\psi]\!]_M \cap [w]_a) \\
M, w \models B_a^n\varphi & \text{iff} & M, v \models \varphi \text{ for all } v \in Min_a^n[w]_a \\
M, w \models \Box_a\varphi & \text{iff} & M, v \models \varphi \text{ for all } v \text{ with } w \succeq_a v
\end{array}$$

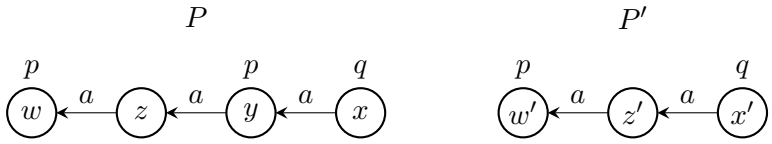

Figure 4: A plausibility model P and its bisimulation contraction P'.

where

$$\begin{array}{lcll} Min_a^0[w]_a & = & Min_{\geq_a}[w]_a & \\ Min_a^{n+1}[w]_a & = & \begin{cases} [w]_a & \text{if } Min_a^n[w]_a = [w]_a \\ Min_a^n[w]_a \cup Min_{\geq_a}([w]_a \setminus Min_a^n[w]_a) & \text{otherwise} \end{cases} & \end{array}$$

and where $[\![\varphi]\!]_M = \{w \in W \mid M, w \models \varphi\}$.

We write $M \models \varphi$ (φ is valid on M) to mean that $M, w \models \varphi$ for all $w \in W$.

Definition 8 (Modal equivalence)**.** Consider the language L_P^X, for X a subsequence of CDS. Given are models $M = (W, \geq, V)$ and $M' = (W', \geq', V')$, and $w \in W$ and $w' \in W'$. We say that (M, w) and (M', w') are *modally equivalent* in L_P^X, notation $(M, w) \equiv_P^X (M', w')$, if and only if for all $\varphi \in L_P^X$, $M, w \models \varphi$ if and only if $M', w' \models \varphi$. If P is obvious from the context we may write $(M, w) \equiv^X (M', w')$.

The logic of conditional belief

The logic L^C of conditional belief appears in [30, 6, 32, 8], where particularly the latter two are foundational for dynamic belief revision (older roots are Lewis' counterfactual conditionals [25]). An axiomatisation is found in [30]. In this logic, defeasible belief $B_a\varphi$ is definable as $B_a^\top \varphi$, while $K_a\varphi$ is definable as $B_a^{\neg\varphi}\bot$.

Example 5. Consider Figure 1. In the plausibility model M_C we have, for instance: $M_C \models K_a p \to (B_a B_b q \wedge \neg K_a B_b q)$: If a knows p (true in v_1 and v_2), a believes, but does not know, that b believes q. Another example is $M_C \models B_a^{\neg B_b q} K_b \neg q$: Conditional on b not believing q, a believes that b knows $\neg q$. Only in v_2 does $\neg B_b q$ hold; there $K_b \neg q$ holds. A final example is $M_C \models K_a p \to B_a^{\widehat{K}_b q} B_b q$: From v_1 and v_2 (where $K_a p$ holds), formula $\widehat{K}_b q$ only holds in v_1, and conditional to that, the one and only most plausible world v_1 satisfies $B_b q$. We can repeat this exercise in M_L and M_R, as all three models are bisimilar and therefore, as will be proved in the next section, logically equivalent.

The logic of degrees of belief

The logic L^D of degrees of belief, also known as the logic of graded belief, goes back to Grove [16] and Spohn [29], although these could more properly be said to be semantic frameworks to model degrees of belief (there is no relation between the logic of degrees of belief and Fine's logic of graded belief [14] and subsequent works, wherein we count the number of pairs $(v, w) \in R$ between two worlds v and w, or, alternatively, label the accessibility relation with that number). Logics of degrees of belief have seen some popularity in artificial intelligence and AGM style belief revision, see e.g. [34, 35, 22]. Belief revision based on degrees of belief have been proposed by [5, 36]. The typical distinction between conviction (arbitrarily strong

belief) and knowledge, as in [23, 24], is absent in our logic L^D, wherein the strongest form of belief defines knowledge. Reasoning with degrees of belief is often called quantitative, where conditional belief can then be called qualitative. In other communities both are called qualitative, and quantitive epistemic reasoning approaches are in that case those that combine knowledge and probabilities [17]. The zeroth degree of belief $B_a^0\varphi$ defines defeasible belief $B_a\varphi$. How Spohn's work relates to dynamic belief revision as in [6] is discussed in detail in [37]. There have also been various proposals combining knowledge and belief ($B_a^\top\varphi$ or $B_a^0\varphi$) in a single framework, without considering either conditional or degrees of belief, where the dynamics are temporal modalities, see [20, 19, 15]. For purposes of further discussions and the proofs in Section 4.2 we define belief layers as follows:

Definition 9 (Belief Layers)**.** Let $M = (W, \geq, V)$. For $w \in W$, $a \in A$ and $n \in \mathbb{N}$, the nth (belief) layer of w for a is defined as $E_a^n[w]_a = Min_{\geq_a}([w]_a \setminus Min_a^{n-1}[w]_a)$, where we use the special case $Min_a^{-1}[w]_a = \emptyset$.

This immediately gives the following lemma:

Lemma 3. *For $M = (W, \geq, V)$, $w \in W$, $a \in A$ and $n \in \mathbb{N}$, we have $Min_a^n[w]_a = E_a^n[w]_a \cup Min_a^{n-1}[w]_a$. For n such that $Min_a^n[w]_a = [w]_a$ we have $E_a^{n+1}[w]_a = \emptyset$. We name the smallest such n the* maximum degree *(for a at w). If n is the maximum degree for a at w, we have $M, w \models K_a\varphi \leftrightarrow B_a^n\varphi$.*

In [5, 36, 22] different layers can contain bisimilar worlds. In our approach they cannot, because we define belief layers on the normal plausibility relation. Unlike [29] our semantics does not allow empty layers in between non-empty layers. If $E_a^n[w]_a \neq \emptyset$ and $E_a^{n+2}[w]_a \neq \emptyset$, then $E_a^{n+1}[w]_a \neq \emptyset$. Layers above the maximum degree will be empty, i.e. if there is a maximum degree n for a at w, as there will always be in our (pre)image-finite models, then for all $k > n$, we have $E_a^k[w]_a = \emptyset$.

Example 6. In Figure 1, we have that $M_C \models B_a^0 B_b^0 q$ but not $M_C \models B_a^1 B_b^0 q$. The maximum degree of belief for a in M_C is at either v_1 and v_2, where it is 1, so $M_C \models K_a\varphi \leftrightarrow B_a^1\varphi$. This is also true in the other two models. Consider now the models P and P' in Figure 4 and an alternative definition of B_a^n not using $\succeq_a$ but $\geq_a$ (as in [5, 36, 22, 8]). In the $\geq_a$-semantics we have $P \models B_a^2\neg q$, as q is false in $\{y, z, w\}$. Only when we reach the third degree of belief does q become uncertain: $P \not\models B_a^3\neg q$. With $\succeq_a$-semantics, 2 is the maximum degree so $P \not\models B_a^2\neg q$. This can be seen in the bisimilar model P', where $P' \not\models B_a^2\neg q$.

The logic of safe belief

The logic L^S of safe belief goes back to Stalnaker [30] and has been progressed by Baltag and Smets (for example, how it relates to conditional belief and knowledge) in [8], which also gives a detailed literature review involving the roots of conditional belief, degrees of belief, and safe belief. An agent has *safe belief* in a formula φ iff the agent will continue to believe φ no matter what *true* information conditions its belief, i.e. $M, w \models \Box_a\varphi$ iff $M, w \models B_a^\psi\varphi$ for all $\Box$-free ψ s.t. $M, w \models \psi$. In [8] safe belief is defined as $M, w \models \Box_a\varphi$ iff $M, v \models \varphi$ for all v s.t. $w \geq_a v$. For both [30] and [8] true information are subsets of the domain containing the actual world. When this is what true information is, there is a correspondence between the two definitions, as indeed noted by Baltag and Smets. The complications of this choice are addressed in detail in [13]. For us, there is not a correspondence between the two definitions,

because we can only condition on modally definable subsets. When we, as we do, define safe belief using $\succeq_a$, this correspondence is reestablished.

Example 7. Consider for a final time the models of Figure 1. We have $M_C, v_1 \models \Box_a \widehat{K}_b q$, whereas $M_C, v_2 \not\models \Box_a \widehat{K}_b q$. Now consider M_L and the $\geq_a$-version of safe belief for which we have $M_L, w_3 \not\models \Box_a \widehat{K}_b q$. For [30, 8] this is as it should be: For the subset $\{w_2, w_3\}$ (which includes the actual world w_3 as required) we have $Min_a(\{w_2, w_3\} \cap [w_3]_a) = \{w_2\}$ where $M_L, w_2 \not\models \widehat{K}_b q$. Using the $\succeq_a$-version of safe belief, we have $M_L, w_3 \models \Box_a \widehat{K}_b q$. For us, this is as it should be: As our conditional belief picks using $[\![\psi]\!] \cap [w]_a$, any set containing w_3 must include the modally equivalent world w_1. This corresponds to first normalising M_L to get M_R. In *that* model, u_2 is strictly less plausible than u_3.

The semantics we propose for degrees of belief and safe belief are non-standard. Still, as we show in the following, these non-standard semantics and the standard semantics for conditional belief are all bisimulation invariant. This makes the results in Section 5 showing a non-trivial expressivity hierarchy between these logics even more remarkable.

4 Bisimulation characterisation for L^C, L^D and L^S

4.1 Bisimulation correspondence for conditional belief

In the following we prove that for the language L^C bisimilarity implies modal equivalence and vice versa. This shows that our notion of bisimulation is proper for the language and models at hand. The proof of Proposition 4 below is essentially a standard proof of bisimilarity implying modal equivalence: modal equivalence is proved by induction on the structure of the formula, where in the induction step the back and forth conditions of bisimilarity are applied to the induction hypothesis. However, the induction case of conditional belief formulas $B_a^\gamma \psi$ is a bit more involved than for standard modalities. Additional work is needed to ensure that when applying the back and forth conditions we produce a world which is among the minimal γ-worlds.

Proposition 4. *Bisimilarity implies modal equivalence for L^C.*

Proof. Assume $(M_1, w_1) \leftrightarrows (M_2, w_2)$. Then, by definition, there exists an autobisimulation R on the disjoint union of M_1 and M_2 with $(w_1, w_2) \in R$. Let $M = (W, \geq, V)$ denote the disjoint union of M_1 and M_2. We then need to prove that (M, w_1) and (M, w_2) are modally equivalent in L_C. We will show that for all φ in L_C, for all $(w, w') \in R$, if $M, w \models \varphi$ then $M, w' \models \varphi$. This implies the required (the other direction being symmetric). The proof is by induction on the syntactic complexity of φ. The propositional cases are easy, so we only consider the cases $\varphi = K_a\psi$ and $\varphi = B_a^\gamma \psi$. Consider first $\varphi = K_a\psi$. In this case we assume $M, w \models K_a\psi$, that is, $M, v \models \psi$ for all v with $w \sim_a v$. Let v' be chosen arbitrarily with $w' \sim_a v'$. We need to prove $M, v' \models \psi$. From Lemma 2 we have that $\geq_a^R$ is a well-preorder on each $\sim_a$-equivalence class. Since $w' \sim_a v'$ we hence get that either $w' \geq_a^R v'$ or $v' \geq_a^R w'$. We can assume $w' \geq_a^R v'$, the other case being symmetric. Then since $(w, w') \in R$ and $w' \geq_a^R v'$, [back$_\geq$] gives us a v s.t. $(v, v') \in R$ and $w \geq_a^R v$. Lemma 2 now implies $w \sim_a v$, and hence $M, v \models \psi$. Since $(v, v') \in R$, the induction hypothesis gives us $M, v' \models \psi$, and we are done.

Now consider the case $\varphi = B_a^\gamma \psi$. This case is more involved. Assume $M, w \models B_a^\gamma \psi$, that is, $M, v \models \psi$ for all $v \in Min_a([\![\gamma]\!]_M \cap [w]_a)$. Letting $v' \in Min_a([\![\gamma]\!]_M \cap [w']_a)$, we need to show

$M, v' \models \psi$ (if $Min_a([\![\gamma]\!]_M \cap [w']_a)$ is empty there is nothing to show). We will first find a y in $Min_a([\![\gamma]\!]_M \cap [w]_a)$, then find a y' with $(y, y') \in R$, and only then apply [back$_\geq$] to $y' \geq_a^R v'$ to produce the required v. The point is that our choice of y in $Min_a([\![\gamma]\!]_M \cap [w]_a)$ will ensure that v is in $Min_a([\![\gamma]\!]_M \cap [w]_a)$.

As mentioned, we want to start out choosing a y in $Min_a([\![\gamma]\!]_M \cap [w]_a)$, so we need to ensure that this set is non-empty. By choice of v' we have $v' \in [\![\gamma]\!]_M$ and $v' \sim_a w'$. From $v' \sim_a w'$ we get that $w' \geq_a^R v'$ or $w' \leq_a^R v'$, using Lemma 2. Since also $(w, w') \in R$, we can apply [back$_\geq$] or [back$_\leq$] to get a u such that $(u, v') \in R$ and either $w \geq_a^R u$ or $w \leq_a^R u$. From $(u, v') \in R$ and $v' \in [\![\gamma]\!]_M$, we get $u \in [\![\gamma]\!]_M$, using the induction hypothesis. From the fact that either $w \geq_a^R u$ or $w \leq_a^R u$ we get $w \sim_a u$, using Lemma 2. Hence we have $u \in [\![\gamma]\!]_M \cap [w]_a$. This shows the set $[\![\gamma]\!]_M \cap [w]_a$ to be non-empty. Hence also $Min_a([\![\gamma]\!]_M \cap [w]_a)$ is non-empty, and we are free to choose a y in that set. Since $y \sim_a w$, Lemma 2 gives us that either $y \geq_a^R w$ or $w \geq_a^R y$, so we can apply [forth$_\leq$] or [forth$_\geq$] to find a y' with $(y, y') \in R$ and either $y' \geq_a^R w'$ or $w' \geq_a^R y'$.

Claim 1. $y' \geq_a^R v'$.

Proof of claim 1. We need to prove $Min_a([y']_{R^=} \cap [y']_a) \geq_a Min_a([v']_{R^=} \cap [v']_a)$. We first prove that $[y']_{R^=} \cap [y']_a \subseteq [\![\gamma]\!]_M \cap [w']_a$:

- $[y']_{R^=} \cap [y']_a \subseteq [\![\gamma]\!]_M$: Assume $y'' \in [y']_{R^=} \cap [y']_a$. Then $(y', y'') \in R^=$. Since we also have $(y, y') \in R$, we get $(y, y'') \in R^=$. From $(y, y'') \in R^=$ and $y \in [\![\gamma]\!]_M$ a finite sequence of applications of the induction hypothesis gives us $y'' \in [\![\gamma]\!]_M$.

- $[y']_{R^=} \cap [y']_a \subseteq [w']_a$: Assume $y'' \in [y']_{R^=} \cap [y']_a$. Then $y'' \sim_a y'$. Since we have either $y' \geq_a^R w'$ or $w' \geq_a^R y'$, we must also have $y' \sim_a w'$, by Lemma 2. Hence $y'' \sim_a y' \sim_a w'$ implying $y'' \in [w']_a$.

Since v' is chosen minimal in $[\![\gamma]\!]_M \cap [w']_a$ and $[y']_{R^=} \cap [y']_a \subseteq [\![\gamma]\!]_M \cap [w']_a$ we get $Min_a([y']_{R^=} \cap [y']_a) \geq_a \{v'\} \geq_a Min_a([v']_{R^=} \cap [v']_a)$, as required. This concludes the proof of the claim.

By choice of y' we have $(y, y') \in R$, and by Claim 1 we have $y' \geq_a^R v'$. We can now finally, as promised, apply [back$_\geq$] to these premises to get a v s.t. $(v, v') \in R$ and $y \geq_a^R v$.

Claim 2. $Min_a([v]_{R^=} \cap [v]_a) \subseteq Min_a([\![\gamma]\!]_M \cap [w]_a)$.

Proof of claim 2. Let $x \in Min_a([v]_{R^=} \cap [v]_a)$. We need to prove $x \in Min_a([\![\gamma]\!]_M \cap [w]_a)$. We do this by proving $x \in [\![\gamma]\!]_M$, $x \in [w]_a$ and $\{x\} \leq_a Min_a([\![\gamma]\!]_M \cap [w]_a)$:

- $x \in [\![\gamma]\!]_M$: By choice of x we have $(v, x) \in R^=$. From $(v, x) \in R^=$ and $(v, v') \in R$ we get $(v', x) \in R^=$. From $(v', x) \in R^=$ and $v' \in [\![\gamma]\!]_M$ a finite sequence of applications of the induction hypothesis gives us $x \in [\![\gamma]\!]_M$.

- $x \in [w]_M$: By choice of x we have $x \sim_a v$. Since $y \geq_a^R v$, Lemma 2 implies $v \sim_a y$. By choice of y we have $y \sim_a w$, so in total we get $x \sim_a v \sim_a y \sim_a w$, as required.

- $\{x\} \leq_a Min_a([\![\gamma]\!]_M \cap [w]_a)$:

$$\begin{array}{lll} \{x\} & \leq_a Min_a([v]_{R^=} \cap [v]_a) & \text{by choice of } x \\ & \leq_a Min_a([y]_{R^=} \cap [y]_a) & \text{since } y \geq_a^R v \\ & \leq_a \{y\} & \\ & \leq_a Min_a([\![\gamma]\!]_M \cap [w]_a) & \text{since } y \in Min_a([\![\gamma]\!]_M \cap [w]_a). \end{array}$$

This concludes the proof of the claim.

Now we are finally ready to prove $M, v' \models \psi$. Let $z \in Min_a([v]_{R^=} \cap [v]_a)$. Then $z \in Min_a([\![\gamma]\!]_M \cap [w]_a)$, by Claim 2. Hence $M, z \models \psi$, by assumption. Since $(v, z) \in R^=$ and $(v, v') \in R$ we get $(z, v') \in R^=$, and hence a finite sequence of applications of the induction hypothesis gives us $M, v' \models \psi$. □

We proceed now to show the converse, that modal equivalence with regard to L^C implies bisimulation. The proof has the same structure as the Hennessy-Millner approach, though appropriately modified for our purposes. Given a pair of image-finite models M and M', the standard approach is to construct a relation $R \subseteq D(M) \times D(M')$ s.t. $(w, w') \in R$ if $M, w \equiv M', w'$. Using $\Diamond$-formulas, it is then shown that R fulfils the requirements for being a bisimulation, as such formulas denote what is true at worlds accessible by whatever accessibility relation is used in the model. This means that modally equivalent worlds have modally equivalent successors, which is then used to show that R fulfils the required conditions. For our purposes this will not do, as we only have $\widehat{K}_a$-formulas (i.e. for $\sim_a$). Instead, our equivalent to $\Diamond$-formulas are of the form $\widehat{B}_a^{\psi}\varphi$, each such formula corresponding to accessibility to the most plausible ψ-worlds from all worlds in an equivalence class. What we want are formulas corresponding to specific links between worlds, so we first establish that such formulas exists. We thus have formulas with the same function as $\Diamond$-formulas serve in the standard approach.

Proposition 5. *Modal equivalence with respect to L^C implies bisimilarity.*

Proof. Assume $(M_1, w) \equiv^C (M_2, w')$. We wish to show that $(M_1, w) \leftrightarrows (M_2, w')$. Let $M = M_1 \sqcup M_2$ be the disjoint union of M_1 and M_2. We then need to show that $Q = \{(v, v') \in D(M) \times D(M) \mid M, v \equiv^C M, v'\}$ is an autobisimulation on M. Note that as $\equiv^C$ is an equivalence relation, so is Q. We first show that $\Diamond$-like formulas talking about the $\geq_a^Q$-relations between specific worlds in M exist.

Claim 1. Let w and w' be worlds of the model $M = (W, \geq, V)$ where $w \geq_a^Q w'$. Further let $\varphi \in L^C$ be any formula true in w'. There then exists a formula $\psi \in L^C$ such that $([w]_Q \cup [w']_Q) \cap [w]_a = [\![\psi]\!]_M \cap [w]_a$ and $M, w \models \widehat{B}_a^{\psi}\varphi$.

Proof of Claim 1. If two worlds s and s' are not modally equivalent, there exists some distinguishing formula $\Psi_{s,s'}$ with $M, s \models \Psi_{s,s'}$ and $M, s' \not\models \Psi_{s,s'}$. As $\sim_a$ is image-finite (since both $\geq_a$ and its converse are) the following formula is finite:

$$\Psi_t = \bigwedge \{\Psi_{t,t'} \mid t \sim_a t' \wedge (t, t') \notin Q\}$$

The formula Ψ_t distinguishes t from all the worlds in $[t]_a$ that it is not modally equivalent to. If there are no such worlds, Ψ_t is the empty conjunction equivalent to $\top$.

We now return to our two original worlds w and w'. With the assumption that $M, w' \models \varphi$, we show that $\psi = \Psi_w \vee \Psi_{w'}$ is a formula of the kind whose existence we claim. First note that $[\![\Psi_w]\!]_M \cap [w]_a$ contains only those worlds in $[w]_a$ that are modally equivalent to w, exactly as $[w]_Q \cap [w]_a$ does. As $[\![\Psi_w]\!]_M \cup [\![\Psi_{w'}]\!]_M = [\![\Psi_w \vee \Psi_{w'}]\!]_M$ we have $([w]_Q \cup [w']_Q) \cap [w]_a = [\![\Psi_w \vee \Psi_{w'}]\!]_M \cap [w]_a$. To get $M, w \models \widehat{B}_a^{\psi}\varphi$ we need to show that $\exists v \in Min_a([\![\Psi_w \vee \Psi_{w'}]\!]_M \cap [w]_a)$ s.t. $M, v \models \varphi$. Pick an arbitrary $v \in Min_a([w']_Q \cap [w']_a)$. We will now show that this has the required properties.

Let $T = [\![\Psi_w \vee \Psi_{w'}]\!]_M \cap [w]_a$. Since $T = ([w]_Q \cup [w']_Q) \cap [w]_a$, Lemma 1 gives $u \simeq_a^Q w$ or $u \simeq_a^Q w'$ for all $u \in T$. Together with $w \geq_a^Q w'$, this gives $w' \in Min_{\geq_a^Q} T$. Choose $u \in T$ arbitrarily. We then have $u \geq_a^Q w'$ and, by definition, that $Min_a([u]_Q \cap [u]_a) \geq_a Min_a([w']_Q \cap [w']_a)$. By choice of v we can then conclude $\{v\} \leq_a Min_a([w']_Q \cap [w']_a) \leq_a Min_a([u]_Q \cap [u]_a) \leq_a \{u\}$. As u was chosen arbitrarily in T, this shows $v \in Min_a T$. As $v \in [w']_Q$ we have $M, v \equiv^C M, w'$ and by assumption of $M, w' \models \varphi$ that $M, v \models \varphi$. We now have $v \in Min_a([\![\Psi_w \vee \Psi_{w'}]\!]_M \cap [w]_a)$ and $M, v \models \varphi$, completing the proof of the claim.

We now proceed to show that Q fulfils the conditions for being an autobisimulation on M (Definition 2). [atoms] is trivial. Next we show [forth$_\geq$]. Let $(w, w') \in Q$ (i.e. $(M, w) \equiv^C (M, w')$) and $w \geq_a^Q v$. We then have that [forth$_\geq$] is fulfilled if $\exists v' \in W$, s.t. $w' \geq_a^Q v'$ and $(v, v') \in Q$ (i.e. $(M, v) \equiv^C (M, v')$). To this end, we show that assuming for all $v' \in W$, $w' \geq_a^Q v'$ implies $(M, v) \not\equiv^C (M, v')$ leads to a contradiction. This is analogous to how Q is shown to be a bisimulation in standard Hennessy-Millner proofs.

We first show that $\geq_a^Q$ is image-finite. First recall that by assumption on plausibility models, $\geq_a$ is (pre)image finite, that is, both $\geq_a$ and $\leq_a$ are image-finite. It follows that $\sim_a = \geq_a \cup \leq_a$ is image-finite as well. If a relation is image-finite, then so is any subset of the relation. Therefore, as $\geq_a^Q \subseteq \sim_a$, $\geq_a^Q$ must be image-finite. Hence the set of $\geq_a^Q$-successors of w', $S = \{v' \mid w' \geq_a^Q v'\} = \{v'_1, \ldots, v'_n\}$ is also finite. Having assumed that v and none of the v'_is are modally equivalent, we have that there exists a number of distinguishing formulae $\varphi^{v'_i}$, one for each v'_i, such that $M, v \models \varphi^{v'_i}$ and $M, v'_i \not\models \varphi^{v'_i}$. Therefore, $M, v \models \varphi^{v'_1} \wedge \cdots \wedge \varphi^{v'_n}$. For notational ease, let $\varphi = \varphi^{v'_1} \wedge \cdots \wedge \varphi^{v'_n}$.

With $M, v \models \varphi$, Claim 1 gives the existence of a formula ψ, such that $([w]_Q \cup [v]_Q) \cap [w]_a = [\![\psi]\!]_M \cap [w]_a$ and $M, w \models \widehat{B}_a^\psi \varphi$. Due to modal equivalence of w and w', we must have $M, w' \models \widehat{B}_a^\psi \varphi$. This we have iff $\exists u' \in Min_a([\![\psi]\!]_M \cap [w']_a)$, s.t. $M, u' \models \varphi$. By construction of φ, no world v'_i exists such that $w' \geq_a^Q v'_i$ and $M, v'_i \models \varphi$, so we must have $u' >_a^Q w'$. As $u' \in [w']_a$, the definition of $>_a^Q$ gives $Min_a([u']_Q \cap [w']_a) >_a Min_a([w']_Q \cap [w']_a)$, so we get $\exists w'' \in Min_a([w']_Q \cap [w']_a)$ s.t. $u' >_a w''$. As $u' \in Min_a([\![\psi]\!]_M \cap [w']_a)$, we must therefore have $w'' \notin [\![\psi]\!]_M$, and then also $w' \notin [\![\psi]\!]_M$. But as $M, w \models \psi$, we get the sought after contradiction (we initially assumed $(M, w) \equiv^C (M, w')$). We get [back$_\geq$] immediately from Q being an equivalence relation.

Now we get to [forth$_\leq$]. Let $(w, w') \in Q$ and $w \leq_a^Q v$. We have that [forth$_\leq$] is fulfilled if $\exists v' \in W$, s.t. $w' \leq_a^Q v'$ and $(v, v') \in Q$.

Claim 2. There exists a $v' \in [w']_a$ satisfying $(v, v') \in Q$.

Proof of Claim 2. Suppose not. Then v does not have a modally equivalent world in $[w']_a$. Thus there must be some formula φ holding in v that holds nowhere in $[w']_a$. Since $v \in [w]_a$ (using Lemma 2), this implies that $M, w \models \widehat{K}_a \varphi$ and $M, w' \not\models \widehat{K}_a \varphi$. However, this contradicts $(w, w') \in Q$, concluding the proof of the claim.

Let v' be chosen as guaranteed by Claim 2. It now suffices to show $w' \leq_a^Q v'$. From $(v, v') \in Q$ and $v \geq_a^Q w$, [forth$_\geq$] gives a w'' s.t. $v' \geq_a^Q w''$ and $(w, w'') \in Q$. From $v' \geq_a^Q w''$ we get $v' \sim_a w''$, using Lemma 2. Since $v' \in [w']_a$ we further get $w' \sim_a v' \sim_a w''$. Since $(w, w'') \in Q$ and $(w, w') \in Q$ we also get $(w', w'') \in Q$. From $w' \sim_a w''$ and $(w', w'') \in Q$ Lemma 1 gives us $w' \simeq_a^Q w''$. From this and $v' \geq_a^Q w''$ we get $v' \geq_a^Q w'$ and hence $w' \leq_a^Q v'$, as required. This concludes proof of [forth$_\leq$]. As for [back$_\geq$] getting to [back$_\leq$] is easy and left out. □

Theorem 1 (Bisimulation characterisation for L^C)**.** *Let $(M,w),(M',w')$ be plausibility models. Then:*

$$(M,w) \leftrightarrows (M',w') \text{ iff } (M,w) \equiv^C (M',w')$$

Proof. From Proposition 4 and Proposition 5. □

We can now finally give the promised proof of Proposition 1.

Proof of Proposition 1. First note that neither the semantics of L^C nor the proofs of Proposition 4 and 5 rely on the existence of largest autobisimulations. Hence we can use these in proving the proposition. Given a plausibility model $M = (W, \geq, V)$ we define a relation R by $R = \{(w,v) \in W^2 \mid M,w \equiv^C M,v\}$. Since modal equivalence implies bisimilarity (Theorem 5), R is a bisimulation relation (indeed, R is exactly the relation shown to be an autobisimulation in the proof of Proposition 5). Now we have to show that R is the largest autobisimulation. If it was not, there would exist an autobisimulation R' with $R' - R \neq \emptyset$. By definition of R, R' would then contain at least one pair (w,v) with $M,w \not\equiv^C M,v$. However, since bisimilarity implies modal equivalence (Proposition 4), this contradicts R' being an autobisimulation. Hence R must be the largest autobisimulation. It only remains to prove that R is an equivalence relation. However, this is trivial given its definition in terms of modal equivalence. □

4.2 Bisimulation correspondence for degrees of belief

We now show bisimulation characterisation results for the logic of degrees of belief L^D. Let $M = (W, \geq, V)$. Recalling Definition 7, for some world $w \in W$, the set $Min_a^0[w]_a$ contains the minimal worlds with respect to $\succeq_a$ in the $\sim_a$-equivalence class of w. For a given w and a, we refer to the generalised definition $Min_a^n[w]_a$ as (belief) sphere n of w for a. The distinction between Min_a^n and Min_a is important to keep straight! The former *Min*—used to give semantics of the B_a^n modality of L^D—is with respect to the relation $\succeq_a$. The latter *Min* is with respect to $\geq_a$, used to give the semantics of L^C. Dealing as we do in this section with L^D, we first state some necessary observations about the properties of what we call beliefs spheres. When convenient we will simply say that v is in (belief) sphere n for a, understanding that this actually means $v \in Min_a^n[v]_a$.

It follows easily from the definitions, that for any world w, sphere n for a is wholly contained within sphere $n+1$ for a, i.e. $Min_a^n[w]_a \subseteq Min_a^{n+1}[w]_a$.

Lemma 4. *Let $M = (W, \geq, V)$ be a plausibility model and consider $w, v \in W$. If $w \sim_a v$ and $w \notin Min_a^n[w]_a$, we have the following two properties:*

(i) If $v \in Min_a^n[w]_a$, then $w \succ_a v$.

(ii) If $v \in Min_a^{n+1}[w]_a$ then $w \succeq_a v$.

Proof. The truth of (i) easily comes from the definition of Min_a^n. For (ii), we consider two exhaustive cases for v. Either $v \in Min_a^{n+1}[w]_a \setminus Min_a^n[w]_a$ in which case $w \succeq_a v$ follows from $\succeq_a$-minimality, since by assumption $w \in [w]_a \setminus Min_a^n[w]_a$. Otherwise $v \in Min_a^n[w]_a$, and so from $w \notin Min_a^n[w]_a$ and (i) it follows that $w \succ_a v$ and hence also $w \succeq_a v$. □

Now getting to the meat of this section, showing bisimulation correspondence for L^D, we first show that bisimilar worlds belong to spheres of all the same degrees.

Lemma 5. *If* $(M_1, w_1) \leftrightarrows (M_2, w_2)$ *then for all* $n \in \mathbb{N}$, $w_1 \in Min_a^n[w_1]_a$ *iff* $w_2 \in Min_a^n[w_2]_a$.

Proof. Assume $(M_1, w_1) \leftrightarrows (M_2, w_2)$. By definition there exists an autobisimulation R on the disjoint union of M_1 and M_2 with $(w_1, w_2) \in R$. Using Proposition 1, let $R_{\max}$ denote the largest bisimulation on the disjoint union of M_1 and M_2 (so $\succeq_a = \geq_a^{R_{\max}}$). Then $R \subseteq R_{\max}$. We are going to show by contradiction that for any $(w, w') \in R_{\max}$ (which includes (w_1, w_2)) and $n \in \mathbb{N}$, $w \in Min_a^n[w]_a$ iff $w' \in Min_a^n[w']_a$. Suppose that this does not hold. Then there must be some pair of worlds w and w' such that $(w, w') \in R_{\max}$ and either i) $w \in Min_a^n[w]_a$ and $w' \notin Min_a^n[w']_a$, or ii) $w \notin Min_a^n[w]_a$ and $w' \in Min_a^n[w']_a$ for some n. Let n be the smallest natural number for which we have either i) or ii). Because the cases are symmetrical, we deal only with i). Using the alternative definition $Min_a^n[w]_a = E_a^n[w]_a \cup Min_a^{n-1}[w]_a$ we can deal with both $n > 0$ and $n = 0$ simultaneously.

By assumption of the smallest n we have $w \notin Min_a^{n-1}[w]_a$, since $w' \notin Min_a^n[w']_a$ implies $w' \notin Min_a^k[w']_a$ for all $0 \leq k \leq n$ (so we could otherwise have chosen a smaller n). Therefore $w \in E_a^n[w]_a$ and $w' \notin E_a^n[w']_a$. Because $w' \in [w'] \setminus Min_a^n[w']_a$, we know that n is not the maximum degree, so there must be *some* world $v' \in E_a^n[w']_a$ which by definition means that $v' \notin Min_a^{n-1}[w']_a$. With $v' \in E_a^n[w']_a \subseteq Min_a^n[w']_a$ and and $w' \notin Min_a^n[w']_a$, Lemma 4 gives $w' \succ_a v'$, i.e. $w' \succeq_a v'$ and $v' \not\succeq_a w'$. By [back$_\geq$] there is a v s.t. $w \succeq_a v$ and $(v, v') \in R_{\max}$. Because $v' \notin Min_a^{n-1}[w']_a$ we cannot have $v \in Min_a^{n-1}[w]_a$, as we could then again have chosen a smaller n making either i) or ii) true. Thus $v \in [w]_a \setminus Min_a^{n-1}[w]_a$. As $w \in Min_a^n[w]_a$, Lemma 4 gives $v \succeq_a w$, so by [forth$_\geq$] there is a u' s.t. $v' \succeq_a u'$ and $(w, u') \in R_{\max}$.

With $(w, w') \in R_{\max}$ and $(w, u') \in R_{\max}$, we have $(w', u') \in R_{\max}$. As $w' \sim_a u'$ (we have $w' \succeq_a v'$ and $v' \succeq_a u'$), Lemma 1 gives $w' \simeq_a^{R_{\max}} u'$, i.e. $w' \succeq_a u'$ and $w' \preceq_a u'$. As $w' \notin Min_a^n[w']_a$, we then have $u' \notin Min_a^n[w']_a$. As $u' \notin Min_a^n[w']_a$ while $v' \in E_a^n[w']_a \subseteq Min_a^n[w']_a$, Lemma 4 then gives $u' \succ_a v'$. But this contradicts $v' \succeq_a u'$, concluding the proof. □

Proposition 6. *Bisimilarity implies modal equivalence for* L^D.

Proof. Assume $(M_1, w_1) \leftrightarrows (M_2, w_2)$. Let $M = (W, \geq, V)$ denote the disjoint union of M_1 and M_2. Then there exists an autobisimulation R on M with $(w_1, w_2) \in R$. Using Proposition 1, let $R_{\max}$ denote the largest autobisimulation on M. Then $R \subseteq R_{\max}$. We need to prove that $(M, w_1) \equiv^D (M, w_2)$.

We will show that for all $(w, w') \in R_{\max}$, for all $\varphi \in L^D$, $M, w \models \varphi$ iff $M, w' \models \varphi$ (which then also means that it holds for all $(w, w') \in R$). We proceed by induction on the syntactic complexity of φ. The propositional and knowledge cases are already covered by Proposition 4, so we only go for $\varphi = B_a^n \psi$.

Assume $M, w \models B_a^n \psi$. We need to prove that $M, w' \models B_a^n \psi$, that is $M, v' \models \psi$ for all $v' \in Min_a^n[w']_a$. Picking an arbitrary $v' \in Min_a^n[w']_a$, we have $[w']_a = [v']_a$ from Lemma 2, and either $w' \succeq_a v'$ or $w' \preceq_a v'$ (so we also have $v' \in Min_a^n[v']_a$). Using [back$_\geq$] or [back$_\leq$] as appropriate, we get that there is a v such that $w \succeq_a v$ or $w \preceq_a v$, and $(v, v') \in R_{\max}$. From this, $v' \in Min_a^n[v']_a$, and Lemma 5 we get $v \in Min_a^n[v]_a$, allowing us to conclude $v \in Min_a^n[w]_a$ from $[w]_a = [v]_a$. With the original assumption of $M, w \models B_a^n \psi$ we get $M, v \models \psi$. As $(v, v') \in R_{\max}$, the induction hypothesis gives $M, v' \models \psi$. As v' was chosen arbitrarily in $Min_a^n[w']_a$ this gives $M, w' \models B_a^n \psi$. Showing that $M, w' \models B_a^n \psi$ implies $M, w \models B_a^n \psi$ is completely symmetrical and therefore left out. □

We now get to showing that modal equivalence for the language of degrees of belief implies bisimilarity. Trouble is, that the B_a^n modality uses the largest autobisimulation for

deriving the relation $\succeq_a$. This makes it difficult to go the Hennessy-Millner way of showing by contradiction that the modal equivalence relation Q is an autobisimulation.

Instead, we establish that modal equivalence for L^D implies modal equivalence for L^C. We go about this by way of a model and world dependent translation of L^C formulas into L^D formulas (Definition 10). This translation has two properties. First, the translated formula is true at M, w iff the untranslated formula is (Lemma 7)—a quite uncontroversial property. More precisely, letting $M = (W, \geq, R)$ be a plausibility model, then for any $w \in W$, $\gamma \in L^C$ where $\sigma_{M,w}(\gamma)$ is the translation at M, w: $M, w \models \gamma \Leftrightarrow M, w \models \sigma_{M,w}(\gamma)$. Assume further that we have some M', w' such that $(M, w) \equiv^D (M', w')$. As $\sigma_{M,w}(\gamma)$ is a formula of L^D we can conclude $M', w' \models \sigma_{M,w}(\gamma)$. So in all we get that

$$M, w \models \gamma \Leftrightarrow M, w \models \sigma_{M,w}(\gamma) \Leftrightarrow M', w' \models \sigma_{M,w}(\gamma) \qquad (*)$$

The second property is that the translation of γ is the same for worlds modally equivalent for L^D (Lemma 8): If $(M, w) \equiv^D (M', w')$ then $\sigma_{M,w}(\gamma) = \sigma_{M',w'}(\gamma)$. This then gives

$$M', w' \models \sigma_{M,w}(\gamma) \Leftrightarrow M', w' \models \sigma_{M',w'}(\gamma) \Leftrightarrow M', w' \models \gamma \qquad (**)$$

Combining (*) and (**) gives that if $(M, w) \equiv^D (M', w')$ then $M, w \models \gamma$ iff $M', w' \models \gamma$ for any $\gamma \in L^C$, i.e. that $(M, w) \equiv^C (M', w')$. As shown in the previous section, modal equivalence for L^C implies bisimilarity (Proposition 5), and we can therefore finally conclude that modal equivalence for L^D implies bisimilarity (Proposition7).

Lemma 6. *For a plausibility model M, a world $w \in D(M)$, agent $a \in A$ and a formula ψ of L^C, if $[\![\psi]\!]_M \cap [w]_a \neq \emptyset$, there is a unique natural number k for which $Min_a([\![\psi]\!]_M \cap [w]_a) \subseteq E_a^k[w]_a$ $(= Min_{\succeq_a}([w]_a \setminus Min_a^{k-1}[w]_a))$.*

Proof. Let $S = [\![\psi]\!]_M \cap [w]_a$. We first show that all worlds in $Min_a\, S$ are equiplausible with respect to $\succeq_a$.

Take any two worlds $v_1, v_2 \in Min_a\, S$. We wish to show $v_1 \simeq_a^R v_2$, i.e. $Min_a([v_1]_R \cap [v_1]_a) \simeq_a Min_a([v_2]_R \cap [v_2]_a)$, where R is the largest autobisimulation on M. With Proposition4 (bisimilarity implies modal equivalence for L^C) and for $i = 1, 2$, we have that $[v_i]_R \subseteq [\![\psi]\!]_M$. Hence $[v_i]_R \cap [v_i]_a = [v_i]_R \cap [w]_a \subseteq [\![\psi]\!]_M \cap [w]_a = S$. With $v_i \in Min_a\, S$ and $v_i \in [v_i]_R \cap [v_i]_a \subseteq S$, we have $v_i \in Min_a([v_i]_R \cap [v_i]_a)$ (if an element of a set A is minimal in a set $B \supseteq A$, then it is also minimal in A). From this we can conclude that $Min_a([v_i]_R \cap [v_i]_a) \simeq_a \{v_i\}$. Since $v_1 \simeq_a v_2$ we get $Min_a([v_1]_R \cap [v_1]_a) \simeq_a \{v_1\} \simeq_a \{v_2\} \simeq_a Min_a([v_2]_R \cap [v_2]_a)$, concluding the proof that $v_1 \simeq_a^R v_2$.

Due to (pre)image-finiteness of M, $[w]_a$ is finite. This means that for any $v \in [w]_a$ there is a unique natural number k for which $v \in E_a^k[w]_a$. As all worlds in $Min_a\, S$ are $\succeq_a$-equiplausible, we have that $Min_a\, S \subseteq E_a^k[w]_a$ for some unique k. □

Having established that if $[\![\psi]\!]_M \cap [w]_a \neq \emptyset$ then there does indeed exist a unique k st. $Min_a([\![\psi]\!]_M \cap [w]_a) \subseteq E_a^k[w]_a$, we have that the following translation is well-defined.

Definition 10 (Translation $\sigma_{M,w}$)**.** Let $M = (W, \geq, V)$ be a plausibility model and $\gamma \in L^C$ be given. We write $\sigma_{M,w}(\gamma)$ for the *translation* of γ at M, w into a formula of L^D defined as follows:

$$\begin{aligned} \sigma_{M,w}(p) &= p \\ \sigma_{M,w}(\neg\varphi) &= \neg\sigma_{M,w}(\varphi) \\ \sigma_{M,w}(\varphi_1 \wedge \varphi_2) &= \sigma_{M,w}(\varphi_1) \wedge \sigma_{M,w}(\varphi_2) \end{aligned}$$

$$\sigma_{M,w}(B_a^\psi\varphi) = \begin{cases} B_a^k \bigvee\{\sigma_{M,v}(\psi \to \varphi) \mid v \in [w]_a\} \wedge \widehat{B}_a^k \bigvee\{\sigma_{M,v}(\psi) \mid v \in [w]_a\} & \text{if } [\![\psi]\!]_M \cap [w]_a \neq \emptyset \\ K_a \bigvee\{\sigma_{M,v}(\neg\psi) \mid v \in [w]_a\} & \text{if } [\![\psi]\!]_M \cap [w]_a = \emptyset \end{cases}$$

where k is the natural number such that $Min_a([\![\psi]\!]_M \cap [w]_a) \subseteq E_a^k[w]_a$. As $K_a\varphi$ is definable in L^C as $B_a^{\neg\varphi}\bot$, we need no $K_a\varphi$-case in the translation.

We need (pre)image-finiteness of M because the translation of $\sigma_{M,w}(B_a^\psi\varphi)$ is based on either $[w]_a$ or $Min_a([\![\psi]\!]_M \cap [w]_a)$. For $\sigma_{M,w}(B_a^\psi\varphi)$ to be finite, we need finiteness of $[w]_a$.

We now get to showing the first of the promised properties of the translation, namely that the translated formula is true at M, w iff the untranslated formula is.

Lemma 7. *Given a plausibility model $M = (W, \geq, V)$ and $\gamma \in L^C$ we have $M, w \models \gamma$ iff $M, w \models \sigma_{M,w}(\gamma)$ for all $w \in W$.*

Proof. We show both directions by induction on the modal depth of γ. For the base case of a modal depth of 0, we have $\sigma_{M,w}(\gamma) = \gamma$ easily, giving $M, w \models \gamma$ iff $M, w \models \sigma_{M,w}(\gamma)$. The p-, $\neg$-, $\wedge$-cases being quite easy, we deal only with $\gamma = B_a^\psi\varphi$ in the induction step. For that case there are to subcases; whether $\sigma_{M,w}(\gamma)$ is a K_a-formula or not.

($\Rightarrow$) : $M, w \models \gamma$ implies $M, w \models \sigma_{M,w}(\gamma)$.

Take first the case $[\![\psi]\!]_M \cap [w]_a = \emptyset$ where $\sigma_{M,w}(B_a^\psi\varphi) = K_a \bigvee\{\sigma_{M,v}(\neg\psi) \mid v \in [w]_a\}$. If $[\![\psi]\!]_M \cap [w]_a = \emptyset$, then $M, v \models \neg\psi$ for all $v \in [w]_a$. Applying the induction hypothesis gives $M, v \models \sigma_{M,v}(\neg\psi)$ for all $v \in [w]_a$. Then we also have $M, u \models \bigvee\{\sigma_{M,v}(\neg\psi) \mid v \in [w]_a\}$ for all $u \in [w]_a$ and finally $M, w \models K_a \bigvee\{\sigma_{M,v}(\neg\psi) \mid v \in [w]_a\}$.

Now take the case $[\![\psi]\!]_M \cap [w]_a \neq \emptyset$. Letting $S = Min_a([\![\psi]\!]_M \cap [w]_a)$ and k be chosen as in the translation, i.e. such that $S \subseteq E_a^k[w]_a$, we wish to prove that $M, w \models B_a^\psi\varphi$ implies $M, w \models B_a^k \bigvee\{\sigma_{M,v}(\psi \to \varphi) \mid v \in [w]_a\} \wedge \widehat{B}_a^k \bigvee\{\sigma_{M,v}(\psi) \mid v \in [w]_a\}$. We first show $M, w \models \widehat{B}_a^k \bigvee\{\sigma_{M,v}(\psi) \mid v \in [w]_a\}$. Because $M, v \models \psi$ for all $v \in S$, the induction hypothesis gives $M, v \models \sigma_{M,v}(\psi)$ for all $v \in S$. From this we can conclude $M, u \models \bigvee\{\sigma_{M,v}(\psi) \mid v \in S\}$ for all $u \in S$, and thus also $M, u \models \bigvee\{\sigma_{M,v}(\psi) \mid v \in [w]_a\}$ for all $u \in S$. From Lemma 6 we have $S \subseteq Min_a^k[w]_a$, so $M, u \models \bigvee\{\sigma_{M,v}(\psi) \mid v \in [w]_a\}$ for some $u \in Min_a^k[w]_a$. This gives $M, w \models \widehat{B}_a^k \bigvee\{\sigma_{M,v}(\psi) \mid v \in [w]_a\}$. Next is $M, w \models B_a^k \bigvee\{\sigma_{M,v}(\psi \to \varphi) \mid v \in [w]_a\}$.

Claim. If $M, w \models B_a^\psi\varphi$, then for all $v \in E_a^k[w]_a \cap [\![\psi]\!]_M$, $M, v \models \varphi$.

Proof of claim. We show the claim by contradiction, assuming that at least one world in $E_a^k[w]_a \cap [\![\psi]\!]_M$ is a $\neg\varphi$-world. Let v be this $\psi \wedge \neg\varphi$-world. As $v \in E_a^k[w]_a$, we have $\{v\} \simeq_a^{R_{\max}} E_a^k[w]_a \simeq_a^{R_{\max}} S$, and specifically that $\forall s \in S : v \simeq_a^{R_{\max}} s$. This means $\forall s \in S : Min([v]_{R_{\max}} \cap [v]_a) \simeq_a Min_a([s]_{R_{\max}} \cap [s]_a)$. Because $\forall s \in S : Min_a([s]_{R_{\max}} \cap [s]_a) \simeq_a S$, we have $Min([v]_{R_{\max}} \cap [v]_a) \simeq_a S$ and thus some $v' \in Min([v]_{R_{\max}} \cap [v]_a)$ such that $\{v'\} \simeq_a S$. Combining $v' \in [v]_{R_{\max}}$ with Theorem 1 gives $M, v \equiv^C M, v'$ and thus that $M, v' \models \psi \wedge \neg\varphi$. Putting $v' \in [\![\psi]\!]_M$ together with $\{v'\} \simeq_a S$, means that $v' \in S$. As $M, v' \models \neg\varphi$, we have a contradiction of $M, w \models B_a^\psi\varphi$, concluding the proof of the claim.

With $M, w \models B_a^\psi\varphi$, we now have $M, v \models \varphi$ for all $v \in E_a^k[w]_a \cap [\![\psi]\!]_M$, and thus $M, v \models \psi \to \varphi$ for all $v \in E_a^k[w]_a$. Lemma 6 gives $S \subseteq E_a^k[w]_a$, and by definition we have $E_a^k[w]_a \cap Min_a^{k-1}[w]_a = \emptyset$, that is, there are no ψ-worlds below layer k, so $M, v \models \psi \to \varphi$ for all $v \in Min_a^k[w]_a$. Using the induction hypothesis gives $M, v \models \sigma_{M,v}(\psi \to \varphi)$ for all $v \in Min_a^k[w]_a$ and therefore $M, w \models B_a^k \bigvee\{\sigma_{M,v}(\psi \to \varphi) \mid v \in [w]_a\}$, finalising left-to-right direction of the proof.

(⇐) : $M, w \models \sigma_{M,w}(\gamma)$ *implies* $M, w \models \gamma$.

We show the stronger claim that $M, w \models \sigma_{M,w'}(\gamma)$ for some $w' \in D(M)$ implies $M, w \models \gamma$. Let $\gamma = B_a^\psi \varphi$ and suppose that $M, w \models \sigma_{M,w'}(\gamma)$ for some $w' \in D(M)$. We then need to show $M, w \models B_a^\psi \varphi$. First take the case where $[\![\psi]\!]_M \cap [w']_a = \emptyset$. Then $\sigma_{M,w'}(B_a^\psi \varphi) = K_a \bigvee\{\sigma_{M,v'}(\neg\psi) \mid v' \in [w']_a\}$, i.e. $M, w \models K_a \bigvee\{\sigma_{M,v'}(\neg\psi) \mid v' \in [w']_a\}$. This means that $M, v \models \bigvee\{\sigma_{M,v'}(\neg\psi) \mid v' \in [w']_a\}$ for all $v \in [w]_a$, i.e. for any $v \in [w]_a$ there is a $v' \in [w']_a$ such that $M, v \models \sigma_{M,v'}(\neg\psi)$. Applying the induction hypothesis, we get $M, v \models \neg\psi$ for all $v \in [w]_a$. Thus $[\![\psi]\!]_M \cap [w]_a = \emptyset$ and we trivially have $M, w \models B_a^\psi \varphi$.

Now take the case $[\![\psi]\!]_M \cap [w']_a \neq \emptyset$. Letting $S' = Min_a([\![\psi]\!]_M \cap [w']_a)$ and k' be s.t. $S' \subseteq E_a^{k'}[w']_a$, we have $M, w \models B_a^{k'} \bigvee\{(\sigma_{M,v'}(\psi \to \varphi) \mid v' \in [w']_a\} \wedge \widehat{B}_a^{k'} \bigvee\{\sigma_{M,v'}(\psi) \mid v' \in [w']_a\}$. From $M, w \models B_a^{k'} \bigvee\{\sigma_{M,v'}(\psi \to \varphi) \mid v' \in [w']_a\}$ we have $M, v \models \bigvee\{\sigma_{M,v'}(\psi \to \varphi) \mid v' \in [w']_a\}$ for all $v \in Min_a^{k'}[w]_a$, i.e. for any $v \in [w]_a$ there is a $v' \in [w']_a$ such that $M, v \models \sigma_{M,v'}(\psi \to \varphi)$. Applying the induction hypothesis, we get $M, v \models \psi \to \varphi$ for all $v \in Min_a^{k'}[w]_a$. From $M, w \models \widehat{B}_a^{k'} \bigvee\{\sigma_{M,v'}(\psi) \mid v' \in [w']_a\}$ we have $M, v \models \bigvee\{\sigma_{M,v'}(\psi) \mid v' \in [w']_a\}$ for some $v \in Min_a^{k'}[w]_a$, i.e. there is a $v \in [w]_a$ and a $v' \in [w']_a$ such that $M, v \models \sigma_{M,v'}(\psi)$. Applying the induction hypothesis gets us $M, v \models \psi$. Thus we have $M, w \models B_a^{k'}(\psi \to \varphi) \wedge \widehat{B}_a^{k'}\psi$ (where $\psi, \varphi \in L^C$).

From $M, w \models \widehat{B}_a^{k'}\psi$ we have that $[\![\psi]\!]_M \cap [w]_a \neq \emptyset$, so Lemma 6 gives the existence of a k, s.t. $Min_a([\![\psi]\!]_M \cap [w]_a) \subseteq Min_a^k[w]_a$. We also have from $M, w \models \widehat{B}_a^{k'}\psi$ that $k \leq k'$, so $Min_a([\![\psi]\!]_M \cap [w]_a) \subseteq Min_a^{k'}[w]_a$. With $M, w \models B_a^{k'}(\psi \to \varphi)$ we get $M, v \models \psi \to \varphi$ for all $v \in Min_a([\![\psi]\!]_M \cap [w]_a)$, then $M, v \models \varphi$ for all $v \in Min_a([\![\psi]\!]_M \cap [w]_a)$, and finally $M, w \models B_a^\psi \varphi$. □

We have now gotten to the second of the two promised properties; that the translation is the same for worlds modally equivalent for L^D.

Lemma 8. *Given plausibility models* M *and* M'*, for any* $w \in D(M)$ *and* $w' \in D(M')$*, if* $(M, w) \equiv^D (M', w')$ *then for any formula* $\gamma \in L^C$*,* $\sigma_{M,w}(\gamma) = \sigma_{M',w'}(\gamma)$.

Proof. We show this by another induction on the modal depth of γ. For the base case of modal depth 0 we trivially have $\sigma_{M,w}(\gamma) = \sigma_{M',w'}(\gamma)$.

For the induction step we, as before, only deal with $\gamma = B_a^\psi \varphi$. Note first that every world in $[w]_a$ is modally equivalent to at least one world in $[w']_a$. If that wasn't the case, there would be some L^D-formula φ true somewhere in $[w]_a$ and nowhere in $[w']_a$. Then $M, w \models \widehat{K}_a\varphi$ while $M', w' \not\models \widehat{K}_a\varphi$, contradicting $(M, w) \equiv^D (M', w')$. A completely analogous argument gives that every world in $[w']_a$ is modally equivalent to at least one world in $[w]_a$. Thus $[\![\psi]\!]_M \cap [w]_a = \emptyset$ iff $[\![\psi]\!]'_M \cap [w']_a = \emptyset$. We thus have two cases, either both $\sigma_{M,w}(B_a^\psi \varphi)$ and $\sigma_{M',w'}(B_a^\psi \varphi)$ are K_a-formulas, or both are B_a^k-formulas.

We deal first with the case where both translations are K_a-formulas. Here we have $\sigma_{M,w}(B_a^\psi \varphi) = K_a \bigvee\{\sigma_{M,v}(\neg\varphi) \mid v \in [w]_a\}$ and $\sigma_{M',w'}(B_a^\psi \varphi) = K_a \bigvee\{\sigma_{M',v'}(\neg\varphi) \mid v' \in [w']_a\}$. As already shown, for all $v \in [w]_a$ there is a $v' \in [w']_a$ such that $(M, w) \equiv^D (M', v')$, and vice versa. The induction hypothesis gives $\sigma_{M,v}(\neg\varphi) = \sigma_{M',v'}(\neg\varphi)$ for all these vs and v's. Then $\bigvee\{\sigma_{M,v}(\neg\varphi) \mid v \in [w]_a\} = \bigvee\{\sigma_{M',v'}(\neg\varphi) \mid v' \in [w']_a\}$ and thus $\sigma_{M,w}(B_a^\psi \varphi) = \sigma_{M',w'}(B_a^\psi \varphi)$.

Take now the case where both translations are B_a^k-formulas. A similar argument as above gives $\bigvee\{\sigma_{M,v}(\psi \to \varphi) \mid v \in [w]_a\} = \bigvee\{\sigma_{M',v'}(\psi \to \varphi) \mid v' \in [w']_a\}$ and $\bigvee\{\sigma_{M,v}(\psi) \mid v \in [w]_a\} = \bigvee\{\sigma_{M',v'}(\psi) \mid v' \in [w']_a\}$. Letting k and k' be the indices chosen in the translation

of $\sigma_{M,w}(B_a^{\psi}\varphi)$ and $\sigma_{M',w'}(B_a^{\psi}\varphi)$ respectively, we have $\sigma_{M,w}(B_a^{\psi}\varphi) = \sigma_{M',w'}(B_a^{\psi}\varphi)$ if $k = k'$. Assume towards a contradiction that $k > k'$. Lemma 6 now gives $Min_a([\![\psi]\!]_M \cap [w]_a) \cap Min_a^{k'}[w]_a = \emptyset$, so $M, v \models \neg\psi$ for all $v \in Min_a^{k'}[w]_a$. With Lemma 7 we have $M, v \models \sigma_{M,v}(\neg\psi)$ for all $v \in Min_a^{k'}[w]_a$ and thus also that $M, w \models B_a^{k'} \bigvee\{\sigma_{M,v}(\neg\psi) \mid v \in [w]_a\}$. From Lemma 6 we also have $Min_a([\![\psi]\!]_M \cap [w']_a) \subseteq Min_a^{k'}[w']_a$, so $M', v' \not\models \neg\psi$ for some $v' \in Min_a^{k'}[w']_a$. From here we use Lemma 7 to conclude $M', v' \not\models \sigma_{M',v'}(\neg\psi)$ for some $v' \in Min_a^{k'}[w']_a$ and thus $M', w' \not\models B_a^{k'} \bigvee\{\sigma_{M',v'}(\neg\psi) \mid v' \in [w']_a\}$. By the work done so far, this also means $M', w' \not\models B_a^{k'} \bigvee\{\sigma_{M,v}(\neg\psi) \mid v \in [w]_a\}$ which contradicts $(M, w) \equiv^D (M', w')$. The case when $k' > k$ is completely symmetrical, and the proof if thus concluded. □

Proposition 7. *Modal equivalence for L^D implies bisimilarity.*

Proof. Let $M = (W, \geq, V)$ and $M' = (W', \geq', V')$ be two plausibility models. We first show that if $(M, w) \equiv^D (M', w')$ then $(M, w) \equiv^C (M', w')$. Assume $(M, w) \equiv^D (M', w,')$ and let γ be any formula of L^C.

$$\begin{aligned} M, w \models \gamma &\Leftrightarrow M, w \models \sigma_{M,w}(\gamma) && \text{(Lemma 7)} \\ &\Leftrightarrow M', w' \models \sigma_{M,w}(\gamma) && \text{(by assumption)} \\ &\Leftrightarrow M', w' \models \sigma_{M',w'}(\gamma) && \text{(Lemma 8)} \\ &\Leftrightarrow M', w' \models \gamma && \text{(Lemma 7)} \end{aligned}$$

Putting this together with Theorem 5 (modal equivalence for L^C implies bisimilarity), we have that two worlds which are modally equivalent in L^D are also modally equivalent in L^C and therefore bisimilar. □

Theorem 2 (Bisimulation characterisation for L^D)**.** *Let $(M, w), (M', w')$ be plausibility models. Then:*

$$(M, w) \leftrightarrows (M', w') \text{ iff } (M, w) \equiv^D (M', w')$$

Proof. From Proposition 6 and Proposition 7. □

4.3 Bisimulation correspondence for safe belief

We now show bisimulation characterisation results for the logic of degrees of belief L^S.

Proposition 8. *Bisimilarity implies modal equivalence for L^S.*

Proof. Assume $M_1 \leftrightarrows M_2$. Then there is an autobisimulation R' on the disjoint union $M_1 \sqcup M_2$ with $R' \cap (D(M_1) \times D(M_2)) \neq \emptyset$. Extend R' into the largest autobisimulation R on $M_1 \sqcup M_2$ (using Proposition 1). Define $R_1 = R \cap (D(M_1) \times D(M_1))$ and $R_2 = R \cap (D(M_2) \times D(M_2))$.

Claim. Let $i \in \{1, 2\}$ and $w \in D(M_i)$. Then

(i) R_i is the largest autobisimulation on M_i.

(ii) $Min_a([w]_{R^=} \cap [w]_a) = Min_a([w]_{R_i^=} \cap [w]_a)$.

(iii) For any v, $w \geq_a^R v$ iff $w \geq_a^{R_i} v$.

Proof of claim. To prove (i), let S_i denote the largest autobisimulation on M_i. If we can show $S_i \subseteq R_i$ we are done. Since S_i is an autobisimulation on M_i, it must also be an autobisimulation on $M_1 \sqcup M_2$. Thus, clearly, $S_i \subseteq R$, since R is the largest autobisimulation on $M_1 \sqcup M_2$. Hence, since $S_i \subseteq D(M_i) \times D(M_i)$, we get $S_i = S_i \cap (D(M_i) \times D(M_i)) \subseteq R \cap (D(M_i) \times D(M_i)) = R_i$. This shows $S_i \subseteq R_i$, as required.

We now prove (ii). Since $w \in D(M_i)$ we get $[w]_a \subseteq D(M_i)$. Since $R_i = R \cap (D(M_i) \times D(M_i))$ this implies $[w]_R \cap [w]_a = [w]_{R_i} \cap [w]_a$. Now note that since R is the largest autobisimulation on $M_1 \sqcup M_2$ and R_i is the largest autobisimulation on M_i, we have $R = R^{=}$ and $R_i = R_i^{=}$, by Proposition 1 (the largest autobisimulation is an equivalence relation). Hence from $[w]_R \cap [w]_a = [w]_{R_i} \cap [w]_a$ we can conclude $[w]_{R^{=}} \cap [w]_a = [w]_{R_i^{=}} \cap [w]_a$, and then finally $Min_a([w]_{R^{=}} \cap [w]_a) = Min_a([w]_{R_i^{=}} \cap [w]_a)$.

We now prove (iii). Note that if $w \geq_a^R v$ or $w \geq_a^{R_i}$ then $w \sim_a v$ (by Lemma 2). So in proving $w \geq_a^R v \Leftrightarrow w \geq_a^{R_i} v$ for $w \in D(M_i)$, we can assume that also $v \in D(M_i)$. We then get:

$$\begin{array}{lcll} w \geq_a^R v & \Leftrightarrow & Min_a([w]_{R^{=}} \cap [w]_a) \geq_a Min_a([v]_{R^{=}} \cap [v]_a) & \\ & \Leftrightarrow & Min_a([w]_{R_i^{=}} \cap [w]_a) \geq_a Min_a([v]_{R_i^{=}} \cap [v]_a) & \text{by (ii), since } w, v \in D(M_i) \\ & \Leftrightarrow & w \geq_a^{R_i} v. & \end{array}$$

This completes the proof of the claim.

We will now show that for all φ and all $(w_1, w_2) \in R \cap (D(M_1) \times D(M_2))$, if $M_1, w_1 \models \varphi$ then $M_2, w_2 \models \varphi$ (the other direction being symmetric). The proof is by induction on the syntactic complexity of φ. The propositional and knowledge cases are already covered by Proposition 4, so we only need to consider the case $\varphi = \Box_a \psi$. Hence assume $M_1, w_1 \models \Box_a \psi$ and $(w_1, w_2) \in R \cap (D(M_1) \times D(M_2))$. We need to prove $M_2, w_2 \models \Box_a \psi$. Pick an arbitrary $v_2 \in D(M_2)$ with $w_2 \succeq_a v_2$. If we can show $M_2, v_2 \models \psi$, we are done. By (i), R_2 is the largest autobisimulation on M_2. Hence $w_2 \succeq_a v_2$ by definition means $w_2 \geq_a^{R_2} v_2$. Using (iii), we can from $w_2 \geq_a^{R_2} v_2$ conclude $w_2 \geq_a^R v_2$. Since R is an autobisimulation, we can now apply [back$_\geq$] to $(w_1, w_2) \in R$ and $w_2 \geq_a^R v_2$ to get a v_1 with $w_1 \geq_a^R v_1$ and $(v_1, v_2) \in R$. Using (iii) again we can conclude from $w_1 \geq_a^R v_1$ to $w_1 \geq_a^{R_1} v_1$, since $w_1 \in D(M_1)$. By (i), R_1 is the largest autobisimulation on M_1, so $w_1 \geq_a^{R_1} v_1$ is by definition the same as $w_1 \succeq_a v_1$. Since we have assumed $M_1, w_1 \models \Box_a \psi$, and since $w_1 \succeq_a v_1$, we get $M_1, v_1 \models \psi$. Since $(v_1, v_2) \in R$, the induction hypothesis gives us $M_2, v_2 \models \psi$, and we are done. □

As for the previous logics, the converse also holds, that is, modal equivalence with regard to L^S implies bisimulation. This is going to be proved as follows. First we prove that any conditional belief formula φ_C can be translated into a logically equivalent safe belief formula φ_S. This implies that if two pointed models (M, w) and (M', w') are modally equivalent in L^S, they must also be modally equivalent in L^C: Any formula $\varphi_C \in L^C$ is true in (M, w) *iff* its translation $\varphi_S \in L^S$ is true in (M, w) *iff* φ_S is true in (M', w') *iff* φ_C is true in (M', w'). Now we can reason as follows: If two pointed models (M, w) and (M', w') are modally equivalent in L^S then they are modally equivalent in L^C and hence, by Theorem 5, bisimilar. This is the result we were after. We postpone the full proof until Section 5.1, which is where we provide the translation of conditional belief formulas into safe belief formulas (as part of a systematic investigation of the relations between the different languages and their relative expressivity). Here we only state the result:

Proposition 9. *Modal equivalence for L^S implies bisimilarity.*

Proof. See Section 5.1. □

As for the two previous languages, L^C and L^D, we now get the following bisimulation characterisation result.

Theorem 3 (Bisimulation characterisation for L^S)**.** *Let $(M, w), (M', w')$ be plausibility models. Then:*

$$(M, w) \leftrightarrows (M', w') \text{ iff } (M, w) \equiv^S (M', w')$$

Proof. From Proposition 8 and Proposition 9. □

4.4 Combining characterisation results

By combining Theorems 1, 2 and 3 from the previous subsections we immediately have the following result.

Corollary 1. *Bisimilarity corresponds to modal equivalence in the logics of conditional belief, degrees of belief, and safe belief, and in any logic containing two or all three of these belief modalities; and modal equivalence in one of these logics corresponds to modal equivalence in any other.*

For example, we also have that $(M, w) \leftrightarrows (M', w')$ iff $(M, w) \equiv^{CDS} (M', w')$, or that $(M, w) \leftrightarrows (M', w')$ iff $(M, w) \equiv^{DS} (M', w')$. Also, to be explicit, we now have that

- $(M, w) \equiv^C (M', w')$ iff $(M, w) \equiv^D (M', w')$
- $(M, w) \equiv^C (M', w')$ iff $(M, w) \equiv^S (M', w')$
- $(M, w) \equiv^D (M', w')$ iff $(M, w) \equiv^S (M', w')$

In other words, the information content of a pointed plausibility model is equally well described in any of these logics. This seems to suggest that it does not matter which logic you use to describe the information content of such a model, apart from the usual considerations of succinctness. Still, this is not the case: our logics are not equally expressive. This will now be addressed in the next section.

5 Expressivity

In this section we will determine the expressivity hierarchy of the logics under consideration. Abstractly speaking, expressivity is a yardstick for measuring whether two logics are able to capture the same properties of a class of models. More concretely in our case, we will for instance be interested in determining whether the conditional belief modality can be expressed using the degrees of belief modality (observe that the translation in Section 4.2 depends on a particular model). With such results at hand we can for instance justify the inclusion or exclusion of a modality, and it also sheds light upon the strengths and weaknesses of our doxastic notions. To start things off we now formally introduce the notion of expressivity found in [40].

Definition 11. Let L and L' be two logical languages interpreted on the same class of models.

- For $\varphi \in L$ and $\varphi' \in L'$, we say that φ and φ' are *equivalent* ($\varphi \equiv \varphi'$) iff they are true in the same pointed models of said class.[3]
- L' is *at least as expressive* as L ($L \leqq L'$) iff for every $\varphi \in L$ there is a $\varphi' \in L'$ s.t. $\varphi \equiv \varphi'$.
- L and L' are *equally expressive* ($L \equiv L'$) iff $L \leqq L'$ and $L' \leqq L$.
- L' is *more expressive* than L ($L < L'$) iff $L \leqq L'$ and $L' \not\leqq L$.
- L and L' are *incomparable* ($L \bowtie L'$) iff $L \not\leqq L'$ and $L' \not\leqq L$.

Below we will show several cases where $L \not\leqq L'$; i.e. that L' is *not* at least as expressive as L. Our primary modus operandi (obtained by logically negating $L \leqq L'$) will be to show that there is a $\varphi \in L$, where for any $\varphi' \in L'$ we can find two pointed models $(M, w), (M', w')$ such that

$$M, w \models \varphi, \quad M', w' \not\models \varphi \quad \text{and} \quad (M, w \models \varphi' \Leftrightarrow M', w' \models \varphi')$$

In other words, for some $\varphi \in L$, no matter the choice of $\varphi' \in L'$, there will be models which φ distinguishes but φ' does not, meaning that $\varphi \not\equiv \varphi'$.

Our investigation will be concerned with the 7 distinct languages that are obtained by considering each L^X such that X is a non-empty subsequence of CDS. In Section 5.1 our focus is on safe belief, and in Section 5.2 on degrees of belief. Using these results, we provide in Section 5.3 a full picture of the relative expressivity of each of these logics, for instance showing that we can formulate 5 distinct languages up to equal expressivity. We find this particularly remarkable in light of the fact that our notion of bisimulation is the right fit for all our logics.

5.1 Expressivity of Safe Belief

Our first result, Proposition 10, shows that the conditional belief modality can be expressed in terms of the safe belief modality. Similar results can be found elsewhere in the literature, for instance in [13, Fact 31] and [7]. In fact, the overall idea of reducing the binary conditional belief operator to a unary belief operator goes back to [11] and [21].

Below we prove that the identity found in [13] is also a valid identity in our logics, which is not a given as our semantics differ in essential ways. In particular the semantics of safe belief in [13] is a standard modality for $\geq_a$, whereas our semantics uses the derived relation $\succeq_a$. A more in-depth account of this matter is provided in Section 6. Returning to the matter at hand, we point out that our work in Section 4 actually serves our investigations here, as evident from the crucial role of Proposition 4 in the following proof.

Proposition 10. *Let $\varphi, \psi \in L_C$. Then the formula $B_a^{\psi}\varphi \leftrightarrow (\widehat{K}_a\psi \to \widehat{K}_a(\psi \wedge \Box_a(\psi \to \varphi)))$ is valid.*

Proof. We let $M = (W, \geq, V)$ be any plausibility model with $w \in W$, and further let $\succeq_a$ denote the normal plausibility relation for an agent a in M. We will show that $M, w \models B_a^{\psi}\varphi \leftrightarrow (\widehat{K}_a\psi \to \widehat{K}_a(\psi \wedge \Box_a(\psi \to \varphi)))$. To this end we let $X = Min_a([\![\psi]\!]_M \cap [w]_a)$.

[3] With our usage of $\equiv$ it is clear from context whether we're referring to modal equivalence, formulas or languages.

Immediately we have that if $X = \emptyset$ then no world in $[w]_a$ satisfies ψ, thus trivially yielding both $M, w \models B_a^{\psi}\varphi$ and $M, w \models \widehat{K}_a\psi \rightarrow \widehat{K}_a(\psi \wedge \Box_a(\psi \rightarrow \varphi))$. For the remainder we therefore assume X is non-empty. We now work under the assumption that $M, w \models B_a^{\psi}\varphi$ and show that this implies $M, w \models \widehat{K}_a\psi \rightarrow \widehat{K}_a(\psi \wedge \Box_a(\psi \rightarrow \varphi))$.

Claim 1. Let $x \in X$ be arbitrarily chosen, then $M, x \models \psi \wedge \Box_a(\psi \rightarrow \varphi)$.

Proof of claim 1. From $x \in X$ we have first that $M, x \models \psi \wedge \varphi$ and $w \sim_a x$. Since $M, x \models \psi$ this means we have proven Claim 1 if $M, x \models \Box_a(\psi \rightarrow \varphi)$ can be shown. To that effect, consider any $y \in W$ s.t. $x \succeq_a y$, for which we must prove $M, y \models \psi \rightarrow \varphi$. When $M, y \not\models \psi$ this is immediate, and so we may assume $M, y \models \psi$. Since $x \succeq_a y$ we have $Min_a([x]_{R^=} \cap [x]_a) \geq_a Min_a([y]_{R^=} \cap [y]_a)$ with R being the largest autobisimulation on M. As R is an autobisimulation we have worlds x', y' in M such that $(y, y') \in R$, $x \geq_a x'$ and $x' \geq_a y'$. Applying Proposition 4 and $M, y \models \psi$ it follows that $M, y' \models \psi$. Using $\geq_a$-transitivity we have $x \geq_a y'$ and hence $w \sim_a x \sim_a y'$, allowing the conclusion that $y' \in X$. By assumption this means $M, y' \models \psi \wedge \varphi$, and so applying once more Proposition 4 it follows that $M, y \models \psi \rightarrow \varphi$ thus completing the proof of this claim.

To show $M, w \models \widehat{K}_a\psi \rightarrow \widehat{K}_a(\psi \wedge \Box_a(\psi \rightarrow \varphi))$ we take any $x \in X$, for which we have $w \sim_a x$ by definition of X. Combining this with Claim 1 it follows that $M, w \models \widehat{K}_a(\psi \wedge \Box_a(\psi \rightarrow \varphi))$. Consequently this also means that $M, w \models \widehat{K}_a\psi \rightarrow \widehat{K}_a(\psi \wedge \Box_a(\psi \rightarrow \varphi))$, thus completing the proof of this direction.

For the converse assume now that $M, w \models \widehat{K}_a\psi \rightarrow \widehat{K}_a(\psi \wedge \Box_a(\psi \rightarrow \varphi))$. As $X \neq \emptyset$ there is a world $u \in W$ s.t. $w \sim_a u$ and $M, u \models \psi \wedge \Box_a(\psi \rightarrow \varphi)$. Therefore we have $M, u' \models \psi \rightarrow \varphi$ for all $u \succeq_a u'$.

Claim 2. Let $x \in X$ be arbitrarily chosen, then $M, x \models \varphi$.

Proof of claim 2. Since $x \in X$ we have by definition that $M, x \models \psi$. It is sufficient to prove that $u \succeq_a x$ because this implies $M, x \models \psi \rightarrow \varphi$ and hence $M, x \models \varphi$ as required. To show $u \succeq_a x$ we assume towards a contradiction that $u \not\succeq_a x$. Now let R denote the largest bisimulation on M and consider any $x' \in Min_a([x]_{R^=} \cap [x]_a)$. As $R^=$ and $\sim_a$ are both reflexive, we have $x \geq_a x'$. From $u \not\succeq_a x$ we therefore have a $u' \in Min_a([u]_{R^=} \cap [u]_a)$ s.t. $u' \not\succeq_a x'$, $u \sim_a u'$ and $(u, u') \in R$ (thus also $x' >_a u'$). Since $u' \sim_a u$ and $u \sim_a w$ we have also $u' \sim_a w$, and additionally from $x \geq_a x'$ and $x' >_a u'$ we can conclude that $x \geq_a u'$ and $u' \not\succeq_a x$. Using $M, u \models \psi$ and $(u, u') \in R$ we apply Proposition 4 which implies $M, u' \models \psi$. As $x \in X$, $u' \sim_a w$ and $x \geq_a u'$ it must be the case that $u' \in X$. From $u' \not\succeq_a x$ we also have that $x \notin X$, but this this contradicts our initial assumption that $x \in X$. We therefore have $u \succeq_a x$ and hence that $M, x \models \varphi$ which completes the proof of the claim.

Recalling that $M, w \models B_a^{\psi}\varphi$ iff $M, x \models \varphi$ for all $x \in X$, Claim 2 readily shows this direction, and thereby completes the proof. □

This result shows there is an equivalence-preserving translation from formulas in L^C to formulas in L^S, and so we have the following results.

Corollary 2. *For any $\varphi \in L^C$ there is a formula $\varphi' \in L^S$ s.t. $\varphi \equiv \varphi'$.*

Corollary 3. *$L^C \leqq L^S$, $L^S \equiv L^{CS}$ and $L^{DS} \equiv L^{CDS}$.*

From Corollary 3 we have that any expressivity result for L^S also holds for L^{CS}, and similarly for L^{DS} and L^{CDS}. In other words, the conditional belief modality is superfluous in

terms of expressivity when the safe belief modality is at our disposal. What is more, we can now finally give a full proof of Proposition 9.

Proof of Proposition 9. Let (M, w) and (M', w') be plausibility models which are modally equivalent in L^S. For any $\varphi_C \in L^C$ it follows from Corollary 2 that there is a $\varphi_S \in L^S$ s.t. $\varphi_C \equiv \varphi_S$. Therefore

$$M, w \models \varphi_C \Leftrightarrow M, w \models \varphi_S \overset{\equiv^S}{\Longleftrightarrow} M', w' \models \varphi_S \Leftrightarrow M', w' \models \varphi_C$$

and hence $(M, w) \equiv^C (M', w')$. Using Proposition 5 we can conclude $(M, w) \leftrightarrows (M', w')$ as required. □

We now proceed to show that L^{CD} is not at least as expressive as L^S. In doing so we need only work with $A = \{a\}$, meaning that the result holds even in the single-agent case. This is also true for our results in Section 5.2.

Lemma 9. *Let p, q be distinct symbols in P, and let $M = (W, \geq, V)$ and $M' = (W', \geq', V')$ denote the two plausibility models presented in Figure 5. Then for $P' = P \setminus \{q\}$ we have that $(M, w_3) \equiv^{CD}_{P'} (M', w'_3)$.*

Proof. We prove the stronger result that for any $\varphi \in L^{CD}_{P'}$:

$$\text{for each } i \in \{1, 2, 3\} : (M, w_i \models \varphi) \Leftrightarrow (M', w'_i \models \varphi)$$

We proceed by induction on φ and let $i \in \{1, 2, 3\}$. When φ is a propositional symbol r in P', we have that $r \neq q$ and so $r \in V(w_i)$ iff $r \in V'(w'_i)$, thus completing the base case. Negation and conjunction are readily shown using the induction hypothesis.

For $\varphi = K_a\psi$ we have that $M, w_i \models K_a\psi$ iff $M, v \models \psi$ for all $v \in \{w_1, w_2, w_3\}$, since $[w_i]_a = \{w_1, w_2, w_3\}$. Applying the induction hypothesis to each element this is equivalent to $M', v' \models \psi$ for all $v' \in \{w'_1, w'_2, w'_3\}$ iff $M', w'_i \models K_a\psi$ (as $[w'_i]_a = \{w'_1, w'_2, w'_3\}$), which completes this case. Continuing to consider $\varphi = B^\gamma_a\psi$ we can simplify notation slightly, namely $Min_a([\![\gamma]\!]_M \cap [w_i]_a) = Min_a[\![\gamma]\!]_M$ since $[w_i]_a = W$. The same holds for each world w'_i of M'.

Claim 1. For M and M' we have that $w_i \in Min_a[\![\gamma]\!]_M$ iff $w'_i \in Min_a[\![\gamma]\!]_{M'}$.

Proof of Claim 1. For M we have that $w_3 >_a w_2$ and $w_2 >_a w_1$, and similarly $w'_3 >'_a w'_2$ and $w'_2 >'_a w'_1$ for M'. Thus the claim follows from the argument below.

$$\begin{aligned} & w_i \in Min_a[\![\gamma]\!]_M \Leftrightarrow \\ M, w_i \models \gamma \text{ and there is no } j <_a i \text{ s.t. } & M, w_j \models \gamma \overset{\text{(IH)}}{\Longleftrightarrow} \\ M', w'_i \models \gamma \text{ and there is no } j <_a i \text{ s.t. } & M', w'_j \models \gamma \Leftrightarrow \\ & w'_i \in Min_a[\![\gamma]\!]_{M'} \end{aligned}$$

We now have that $M, w_i \models B^\gamma_a\psi$ iff $M, v \models \psi$ for all $v \in Min_a[\![\gamma]\!]_M$. Applying both the induction hypothesis and Claim 1, we have that this is equivalent to $M, v' \models \psi$ for all $v' \in Min_a[\![\gamma]\!]_{M'}$ iff $M', w'_i \models B^\gamma_a\psi$.

Finally we consider the case of $\varphi = B^n_a\psi$. To this end we note that the union of $\{(w'_1, w'_3)\}$ and the identity relation on W is the largest bisimulation on M' (this relation cannot be

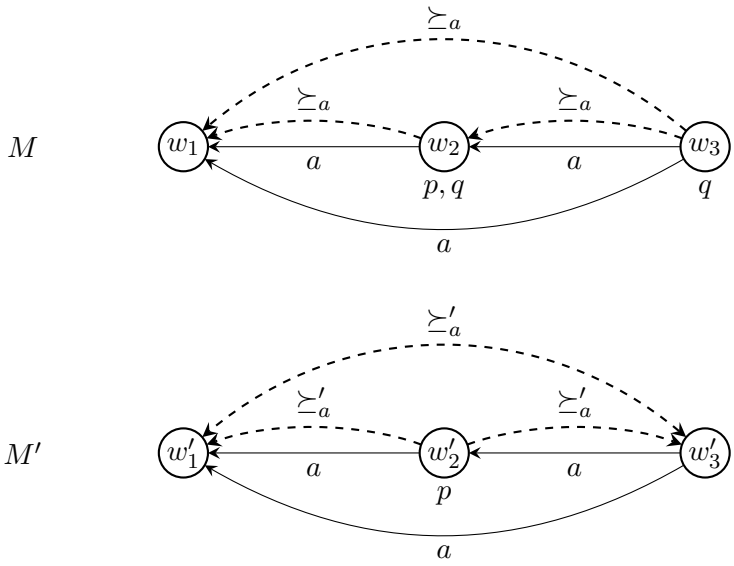

Figure 5: Two single-agent plausibility models and their normal plausibility relations (dashed arrows). As usual reflexive arrows are omitted.

extended and still satify [atoms]). As w'_1 and w'_3 are bisimilar, it follows from Corollary 1 that $M', w'_1 \models \psi$ iff $M', w'_3 \models \psi$ $(*)$.

Claim 2. For $n \in \mathbb{N}$ we have that $M, w \models \psi$ for all $w \in Min^n_a[w_i]$ iff $M', w' \models \psi$ for all $w' \in Min^n_a[w'_i]$.

Proof of Claim 2. We treat three exhaustive cases for n.

- $n = 0$: $M, w \models \psi$ for all $w \in Min^0_a[w_i] \Leftrightarrow M, w_1 \models \psi \overset{\text{(IH)}}{\Longleftrightarrow} M', w'_1 \models \psi \overset{(*)}{\Longleftrightarrow} M', w'_3 \models \psi$. Therefore $M, w \models \psi$ for all $w \in Min^0_a[w_i]$ is equivalent to $M', w' \models \psi$ for all $w' \in \{w'_1, w'_3\}$, and as $Min^0_a[w'_i] = \{w'_1, w'_3\}$ this concludes this case.

- $n = 1$: Since $Min^1_a[w_i] = \{w_1, w_2\}$ we have that $M, w \models \psi$ for all $w \in \{w_1, w_2\} \overset{\text{(IH)}}{\Longleftrightarrow} M', w' \models \psi$ for all $w' \in \{w'_1, w'_2\}$. Using $(*)$ this is equivalent to $M', w' \models \psi$ for all $w' \in \{w'_1, w'_2, w'_3\}$. By this argument and the fact that $Min^1_a[w'_i] = \{w'_1, w'_2, w'_3\}$, we can conclude $M, w \models \psi$ for all $w \in Min^1_a[w_i] \Leftrightarrow M', w' \models \psi$ for all $w' \in Min^1_a[w'_i]$ as required.

- $n \geq 2$: We have that $Min^m_a[w_i] = \{w_1, w_2, w_3\}$ and $Min^m_a[w'_i] = \{w'_1, w'_2, w'_3\}$, hence this is exactly as the case of $\varphi = K\psi$.

We have that $M, w_i \models B^n_a\psi$ iff $M, w \models \psi$ for all $w \in Min^n_a[w_i]$. Applying Claim 2 this is equivalent to $M', w' \models \psi$ for all $w' \in Min^n_a[w'_i]$ iff $M', w'_i \models B^n_a\psi$, thereby completing the final case of the induction step. It follows that $(M, w_3) \equiv^{CD}_{P'} (M', w'_3)$ as required. □

Proposition 11. $L^S \not\leqslant L^{CD}$.

Proof. Consider the formula $\Diamond_a p$ of L^S with $p \in P$, and take some arbitrary formula $\varphi_{CD} \in L^{CD}_P$. As φ_{CD} is finite and P is countably infinite, there will be *some* $q \neq p$ not occurring in φ_{CD}. Letting $P' = P \setminus \{q\}$ this means that $\varphi_{CD} \in L^{CD}_{P'}$. This choice of p and q can always be made, and consequently there also exists models M and M' as given in Figure 5. The largest bisimulation on M is the identity as no two worlds have the same valuation. At the same time $\{(w'_1, w'_1), (w'_1, w'_3), (w'_2, w'_2), (w'_3, w'_1), (w'_3, w'_3)\}$ is the largest bisimulation on M'. This gives rise to the normal plausibility relations $\succeq_a$ (for M) and $\succeq'_a$ (for M') depicted in Figure 5 using dashed edges.

Since $w_3 \succeq_a w_2$ and $M, w_2 \models p$ it follows that $M, w_3 \models \Diamond_a p$. Furthermore we have that the image of w'_3 under $\succeq'_a$ is $\{w'_1, w'_3\}$. This means that there is no $v' \in W'$ s.t. $w'_3 \succeq'_a v'$ and $M', v' \models p$, and consequently $M', w'_3 \not\models \Diamond_a p$. At the same time we have by Lemma 9 that $M, w_3 \models \varphi_{CD}$ iff $M', w'_3 \models \varphi_{CD}$. Therefore using the formula $\Diamond_a p$ of L^S, for any formula of $\varphi_{CD} \in L^{CD}$ there are models which $\Diamond_a p$ distinguishes but φ_{CD} does not, and so $\Diamond_a p \not\equiv \varphi_{CD}$. Consequently we have $L^S \not\leqslant L^{CD}$ as required. □

To further elaborate on this result, what is really being put to use here is the ability of the safe belief modality to (at least in part) talk about propositional symbols that do not occur in a formula. This is an effect of the derived relation $\succeq_a$ depending on the largest bisimulation.

5.2 Expressivity of Degrees of Belief

We have now settled that safe belief is at least as expressive as conditional belief, and further that the combination of the conditional belief modality and the degrees of belief modality does not allow us to express the safe belief modality. A hasty conclusion would be that the safe belief modality is the one modality to rule them all, but this is not so. In fact L^S (equivalent to L^{CS} cf. Corollary 3) falls short when it comes to expressing degrees of belief, which we now continue to prove.

Lemma 10. *Let p, q be distinct symbols in P, and let $M = (W, \geq, V)$ and $M' = (W', \geq', V')$ denote the two plausibility models presented in Figure 6. Then for $P' = P \setminus \{q\}$ we have that $(M, x_1) \equiv^S_{P'} (M', x')$.*

Proof. We will show the following stronger version of this lemma: For $i \in \{1, 2\} : (M, x_i) \equiv^S_{P'} (M', x')$ and $(M, y) \equiv^S_{P'} (M', y')$. We proceed by induction on $\varphi \in L^S_{P'}$, showing that:

$$\text{for } i \in \{1, 2\} : M, x_i \models \varphi \text{ iff } M', x' \models \varphi \qquad \text{and} \qquad M, y \models \varphi \text{ iff } M', y' \models \varphi \tag{1}$$

For the base case we have $\varphi = r$ for some $r \in P \setminus \{q\}$. Because $r \neq q$ it is clear that $r \in V(x_1)$ iff $r \in V'(x')$. Since we also have $V(x_2) = V'(x')$ and $V'(y) = V(y')$ this completes the base case. The cases of negation and conjunction are readily established using the induction hypothesis, and $\varphi = K_a \psi$ is shown just as we did in the proof of Lemma 9. Before proceeding we recall that $A = \{a\}$ and note that for any $w \in W$ we have $[w]_a = \{x_1, x_2, y\}$, as well as $[w']_a = \{x', y'\}$ for any $w' \in W'$. Moreover, the largest bisimulation on M and M' respectively is the identity relation, meaning that $\geq_a = \succeq_a$ and $\geq'_a = \succeq'_a$. For the case of $\varphi = \Box_a \psi$ we can therefore argue as follows.

$$M, x_1 \models \Box_a \psi \Leftrightarrow M, x_1 \models \psi \overset{\text{(IH)}}{\Longleftrightarrow} M', x' \models \psi \Leftrightarrow M', x' \models \Box_a \psi$$

$$M, x_2 \models \Box_a \psi \Leftrightarrow (\forall i \in \{1, 2\} : M, x_i \models \psi) \overset{\text{(IH)}}{\Longleftrightarrow} M', x' \models \psi \Leftrightarrow M', x' \models \Box_a \psi$$

$$M, y \models \Box_a \psi \Leftrightarrow (\forall w \in W : M, w \models \psi) \overset{\text{(IH)}}{\Longleftrightarrow} (\forall w' \in W' : M', w' \models \psi) \Leftrightarrow M', y' \models \Box_a \psi$$

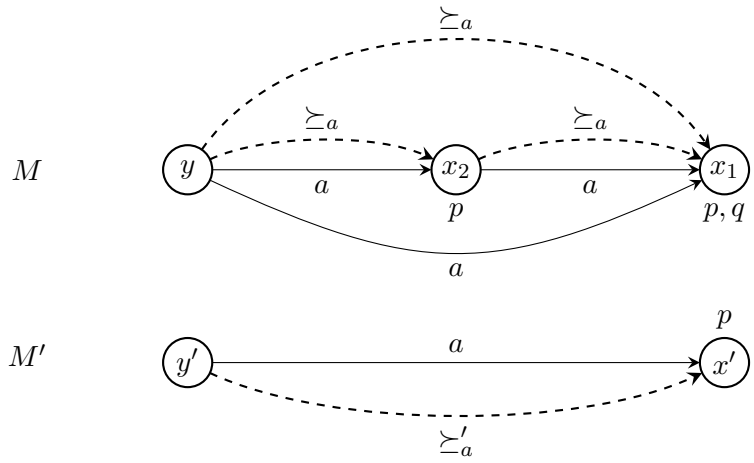

Figure 6: Two single-agent plausibility models and their normal plausibility relations (dashed arrows). As usual reflexive arrows are omitted.

In fact the last line is essentially the case of $K_a\psi$, as the image of y under $\succeq_a$ is W (and W' is the image of y' under $\succeq'_a$). This concludes our proof by induction, shows (1) and allows us to conclude that $(M, x_1) \equiv^S_{P'} (M', x')$. □

Proposition 12. $L^D \not\leqslant L^S$.

Proof. Consider the formula $B^1_a p \in L^D$ with $p \in P$, and additionally take any formula $\varphi_S \in L^S_P$. As φ_S is finite and P is countably infinite, there will be *some* $q \neq p$ which does not occur in φ_S. With $P' = P \setminus \{q\}$ we therefore have $\varphi_S \in L^S_{P'}$. As we can always make such a choice of p and q, this means that there always exists models (M, x_1), (M', x') of the form given in Figure 6.

As in the proof of Lemma 10 the largest bisimulation on M and M' is the identity and so $Min^1_a[x_1]_a = \{x_1, x_2\}$ and $Min^1_a[x']_a = \{x', y'\}$. Consequently $M, x_1 \models B^1_a p$ whereas $M', x' \not\models B^1_a p$. Since $\varphi_S \in L^S_{P'}$ it follows from Lemma 10 that $M, x \models \varphi_S$ iff $M', x' \models \varphi_S$. What this proves is that using the formula $B^1_a p$ of L^D, no matter the choice of formula φ_S of L^S there will be models which $B^1_a p$ distinguishes but φ_S does not, hence $B^1_a p \not\equiv \varphi_S$. From this follows $L^D \not\leqslant L^S$ as required. □

We find that this result is quite surprising. Again it is a consequence of our use of the largest bisimulation when defining our semantics. The purpose of x_1 in model M (which is otherwise identical to M') is to inject an additional belief sphere without adding any factual content from P', as that could allow the safe belief formula φ_S to distinguish x_1 from x_2.

At this point it might seem as if all hope was lost for the conditional belief modality, however our final direct result somewhat rebuilds the reputation of this hard-pressed modality. To this end we define for any $k \in \mathbb{N}$ the language L^{Dk}, which contains every formula of L^D for which if $B^n_a \varphi$ occurs then $n \leq k$. In other words formulas of L^{Dk} talk about belief to at most degree k, which comes in handy as we investigate the relative expressive power of L^D and L^C.

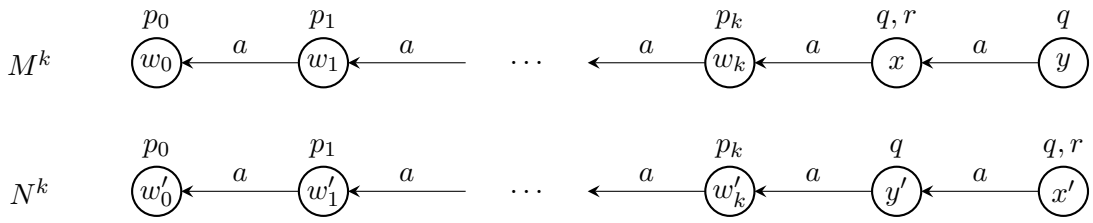

Figure 7: Two single-agent plausibility models. We've omitted reflexive arrows and for the sake of readability also some transitive arrows.

Lemma 11. *Let $k \in \mathbb{N}$ be given, and let (M^k, w_0) and (N^k, w'_0) denote the two plausibility models presented in Figure 7. Then we have that (M^k, w_0) and (N^k, w'_0) are modally equivalent in L^{Dk}.*

Proof. We prove a stronger version of this lemma, namely that $(M^k, w_i) \equiv^{Dk} (N^k, w'_i)$ for $0 \leq i \leq k$, $(M^k, x) \equiv^{Dk} (N^k, x')$ and $(M^k, y) \equiv^{Dk} (N^k, y')$.

Key to this proof is the fact that x (resp. y) has the same valuation as x' (resp. y'), and that x is more plausible than y whereas y' is more plausible than x'. We proceed by induction on $\varphi \in L^{Dk}$. In the base case φ is a propositional symbol, and so as the valuation of each w_i matches that of w'_i $(0 \leq i \leq k)$, x matches x' and y matches y' this completes the base case. The cases of negation and conjunction readily follow using the induction hypothesis, and for $\varphi = K_a\psi$ the argument is essentially that used in the proof of Lemma 9.

Lastly we consider $\varphi = B^j_a\psi$ for any $0 \leq j \leq k$, and recall that this is sufficient as $\varphi \in L^{Dk}_P$. As neither model contains two worlds with the same valuation, the largest autobisimulation on either model is the identity, and so both models are normal. With the epistemic relation of agent a being total, we have for all $w \in W$ that $Min^j_a[w]_a = \{w_0, \ldots, w_j\}$ and similarly for all $w' \in W'$ that $Min^j_a[w']_a = \{w'_0, \ldots, w'_j\}$. We therefore have

$$\begin{aligned}\forall w \in W : M^k, w \models B^j_a\psi \Leftrightarrow \forall v \in \{w_0, \ldots, w_j\} : M^k, v \models \psi \overset{\text{(IH)}}{\Longleftrightarrow} \\ \forall v' \in \{w'_0, \ldots, w'_j\} : N^k, v' \models \psi \Leftrightarrow \ \forall w' \in W' : N^k, w' \models B^j_a\psi\end{aligned}$$

as required. Observe that we can apply the induction hypothesis since $j \leq k$, and that importantly x, y are not in $Min^j_a[w]_a$, and x', y' are not in $Min^j_a[w']_a$. Thus we have shown that $(M^k, w_0) \equiv^{Dk} (N^k, w'_0)$ thereby completing the proof. □

Proposition 13. *$L^C \not\leqslant L^D$.*

Proof. Consider now $B^q_a r$ belonging to L^C and any formula $\varphi_D \in L^D$. Since φ_D is finite we can choose some $k \in \mathbb{N}$ such that $\varphi_D \in L^{Dk}$. Because $p_0, \ldots, p_k, q, r$ are taken from the countably infinite set P, no matter the choice of k there exists pointed plausibility models (M^k, w_0) and (N^k, w'_0) as presented in Figure 7.

To determine the truth of $B^q_a r$ in (M^k, w_0) and (N^k, w'_0) respectively we point out that $[\![q]\!]_{M^k} = \{x, y\}$ and $[\![q]\!]_{N^k} = \{y', x'\}$. Therefore we have that $Min_a([\![q]\!]_{M^k} \cap [w_0]_a) = \{x\}$ and $Min_a([\![q]\!]_{N^k} \cap [w'_0]_a) = \{y'\}$. Since $M^k, x \models r$ and $N^k, y' \not\models r$, it follows $M^k, w_0 \models B^q_a r$ whereas $N^k, w'_0 \not\models B^q_a r$. By Lemma 11 we have that $M^k, w_0 \models \varphi_D$ iff $N^k, w'_0 \models \varphi_D$. With this we have shown that taking the formula $B^q_a r$ of L^C, there are for any $\varphi_D \in L^D$ pointed plausibility models which $B^q_a r$ distinguishes but φ_D does not, thus $B^q_a r \not\equiv \varphi_D$. It follows that $L^C \not\leqslant L^D$ as required. □

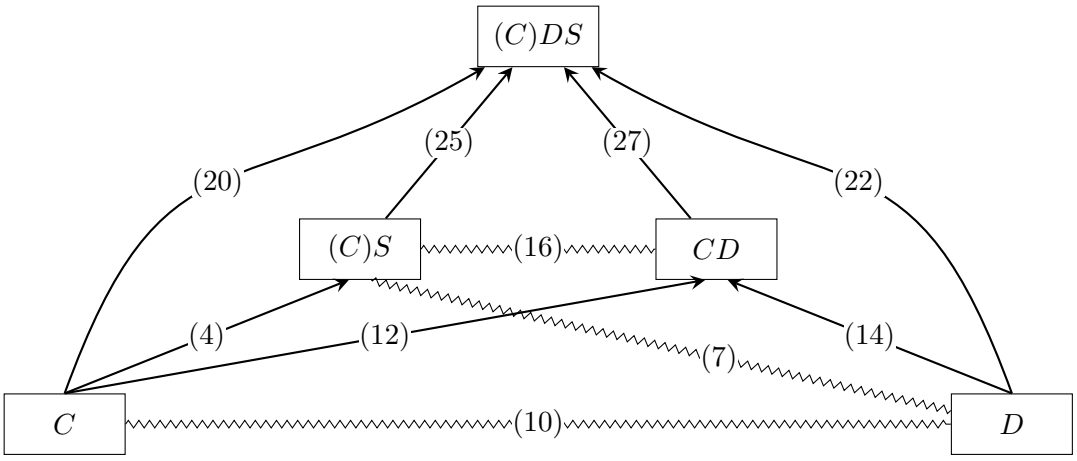

Figure 8: Summary of expressivity results for our logics. An arrow $X \longrightarrow X'$ indicates that $L^{X'}$ is more expressive than L^X. A zig-zag line between X and X' means that L^X and $L^{X'}$ are incomparable. The abbreviation $(C)DS$ means both CDS and DS, and similarly for $C(S)$ indicating both CS and S. Labels on arrows and zig-zag lines signify from where the result is taken in Table 1.

We have now shown that the degrees of belief modality cannot capture the conditional belief modality. What this really showcases is that for $B_a^\psi \varphi$, ψ potentially enables us to talk about worlds of arbitrarily large degree. This sets it apart from the degrees of belief modality, and causes for instance a difference in expressivity.

5.3 Mapping Out the Relative Expressive Power

With the results we have now shown, we are in fact able to determine the relative expressivity of all our languages. To this end we make use of the following facts related to expressivity, where we let L, L' and L'' denote logical languages interpreted on the same class of models:

(a) If L is a sublanguage of L' then $L \leqq L'$.

(b) If $L \leqq L'$ and $L' \leqq L''$ then $L \leqq L''$ (transitivity).

(c) If $L \equiv L'$ then $L \leqq L''$ iff $L' \leqq L''$ (transitivity consequence 1).

(d) If $L \leqq L'$ and $L'' \not\leqq L'$ then $L'' \not\leqq L$ (transitivity consequence 2).

(e) If $L \leqq L'$ and $L \not\leqq L''$ then $L' \not\leqq L''$ (transitivity consequence 3).

Now comes our main result, which shows the relative expressivity between the logic of conditional belief, the logic of degrees of belief and the logic of safe belief.

Theorem 4. $L^C < L^S$, $L^C \bowtie L^D$, $L^D \bowtie L^S$.

Proof. See the derivation of (4), (7) and (10) in Table 1. □

Beyond showing the above theorem, Table 1 fully accounts for the relative expressivity between L^C, L^D, L^S, L^{CD} and L^{DS}. Finally, using Corollary 3 and property (c) we have that any expressivity result for L^S holds for L^{CS} and similarly for L^{DS} and L^{CDS}. A more pleasing presentation of these results is found in Figure 8.

#	Result	Inferred from
(1)	$L^C \leqq L^S$	Corollary 3.
(2)	$L^S \not\leqq L^{CD}$	Proposition 11.
(3)	$L^S \not\leqq L^C$	$L^C \leqq L^{CD}$ from (a), $L^S \not\leqq L^{CD}$ from (2) and applying (d).
(4)	$L^C < L^S$	$L^C \leqq L^S$ from (1), $L^S \not\leqq L^C$ from (3).
(5)	$L^D \not\leqq L^S$	Proposition 12.
(6)	$L^S \not\leqq L^D$	$L^D \leqq L^{CD}$ from (a), $L^S \not\leqq L^{CD}$ from (2) and applying (d).
(7)	$L^D \bowtie L^S$	$L^D \not\leqq L^S$ from (5), $L^S \not\leqq L^D$ from (6).
(8)	$L^C \not\leqq L^D$	Proposition 13.
(9)	$L^D \not\leqq L^C$	$L^C \leqq L^S$ from (1), $L^D \not\leqq L^S$ from (5) and applying (d).
(10)	$L^C \bowtie L^D$	$L^C \not\leqq L^D$ from (8), $L^D \not\leqq L^C$ from (9).
(11)	$L^{CD} \not\leqq L^C$	$L^D \leqq L^{CD}$ from (a), $L^D \not\leqq L^C$ from (9) and applying (e).
(12)	$L^C < L^{CD}$	$L^C \leqq L^{CD}$ from (a), $L^{CD} \not\leqq L^D$ from (13).
(13)	$L^{CD} \not\leqq L^D$	$L^C \leqq L^{CD}$ from (a), $L^C \not\leqq L^D$ from (8) and applying (e).
(14)	$L^D < L^{CD}$	$L^D \leqq L^{CD}$ from (a), $L^{CD} \not\leqq L^D$ from (13).
(15)	$L^{CD} \not\leqq L^S$	$L^D \leqq L^{CD}$ from (a), $L^D \not\leqq L^S$ from (5) and applying (e).
(16)	$L^S \bowtie L^{CD}$	$L^S \not\leqq L^{CD}$ from (2), $L^{CD} \not\leqq L^S$ from (15).
(17)	$L^{CDS} \leqq L^{DS}$	$L^{CDS} \equiv L^{DS}$ from Corollary 3 and Definition 11.
(18)	$L^C \leqq L^{DS}$	$L^C \leqq L^{CDS}$ from (a), $L^{CDS} \leqq L^{DS}$ from (17) and applying (b).
(19)	$L^{DS} \not\leqq L^C$	$L^S \leqq L^{DS}$ from (a), $L^S \not\leqq L^C$ from (3) and applying (e).
(20)	$L^C < L^{DS}$	$L^C \leqq L^{DS}$ from (18), $L^{DS} \not\leqq L^C$ from (19).
(21)	$L^{DS} \not\leqq L^D$	$L^S \leqq L^{DS}$ from (a), $L^S \not\leqq L^D$ from (6) and applying (e).
(22)	$L^D < L^{DS}$	$L^D \leqq L^{DS}$ from (a), $L^{DS} \not\leqq L^D$ from (21).
(23)	$L^{CD} \leqq L^{DS}$	$L^{CD} \leqq L^{CDS}$ from (a), $L^{CDS} \leqq L^{DS}$ from (17) and applying (b).
(24)	$L^{DS} \not\leqq L^S$	$L^{CD} \leqq L^{DS}$ from (23), $L^{CD} \not\leqq L^S$ from (15) and applying (e).
(25)	$L^S < L^{DS}$	$L^S \leqq L^{DS}$ from (a), $L^{DS} \not\leqq L^S$ from (24).
(26)	$L^{CD} \not\leqq L^{DS}$	$L^S \leqq L^{DS}$ from (a), $L^S \not\leqq L^{CD}$ from (2) and applying (e).
(27)	$L^{CD} < L^{DS}$	$L^{CD} \leqq L^{DS}$ from (23), $L^{CD} \not\leqq L^{DS}$ from (26).

Table 1: Derivation of the relative expressivity of our logics. Each of the references (a), (b), (d) and (e) refer to properties stated at the start of Section 5.3. Bold faced numbers are illustrated in Figure 8.

5.4 Reflection on bisimulation characterisation and expressivity

Our bisimulation characterisation results are:

$$(M,w) \underline{\leftrightarrows} (M',w') \quad \text{iff} \quad (M,w) \equiv^C (M',w') \qquad \text{Theorem 1}$$
$$(M,w) \underline{\leftrightarrows} (M',w') \quad \text{iff} \quad (M,w) \equiv^D (M',w') \qquad \text{Theorem 2}$$
$$(M,w) \underline{\leftrightarrows} (M',w') \quad \text{iff} \quad (M,w) \equiv^S (M',w') \qquad \text{Theorem 3}$$

In other words, bisimulation corresponds to modal equivalence in all three logics. Our expressivity results can be summarised as (Theorem 4)

$$\begin{array}{lcl} L^C & < & L^S \\ L^C & \bowtie & L^D \\ L^D & \bowtie & L^S \end{array}$$

The logic of conditional belief is less expressive than the logic of safe belief, the logic of conditional belief and the logic of degrees of belief are incomparable, as are the logic of degrees of belief and the logic of safe belief.

Our results on bisimulation characterisation suggest that, in some sense, the three logics are the same, whereas our results on expressive power suggest that, in another sense, the three logics are different. It is therefore a good moment to explain how to interpret our results.

The bisimulation characterisation result in Corollary 1 says that the information content of a given plausibility model is equally well described in the three logics. Now consider an even more specific case: a finite model; and consider a characteristic formula of that model (these can be shown to exist for plausibility models along the lines of [31, 38]—where we note that we take models, not pointed models). For a model M this gives us, respectively, formulas φ^C_M, φ^D_M, and φ^S_M. Then the bisimulation characterisation results say that φ^C_M, φ^D_M, and φ^S_M are all equivalent. Now a characteristic formula is a very special formula with a unique model (modulo bisimilarity). For other formulas that do not have a singleton denotation (again, modulo bisimilarity) in the class of plausibility models, this equivalence cannot be achieved. That is the expressivity result. For example, given that $L^C < L^S$, there is a safe belief formula that is not equivalent to any conditional belief formula. This formula should then describe a property that has several non-bisimilar models. It is indeed the case that the formula $\Diamond_a p$ used in the proof of Proposition 11 demonstrating $L^C < L^S$ has many models! It is tempting to allow ourselves a simplication and to say that the expressivity hierarchy breaks down if we restrict ourselves to formulas with unique models.[4]

Finally, we must point out that in the publication on single-agent bisimulation [4, p. 285], we posed the following conjecture:

> *In an extended version of the paper we are confident that we will prove that the logics of conditional belief and knowledge, of degrees of belief and knowledge, and both with the addition of safe belief are all expressively equivalent.*

It therefore seems appropriate to note that we have proved our own confident selves resoundingly wrong!

6 Comparison and applications

We compare our bisimulation results to those in Demey's work [13], our expressivity results to those obtained in Baltag and Smets' [7], and finally discuss the relevance of our results for epistemic planning [10].

[4] If we consider infinitary versions of the modalities in our logical languages, in other words, common knowledge and common belief modalities, we preserve the bisimulation characterisation results (for a more refined notion of bisimulation) but it is then to be expected that all three logics become equally expressive (oral communication by Tim French).

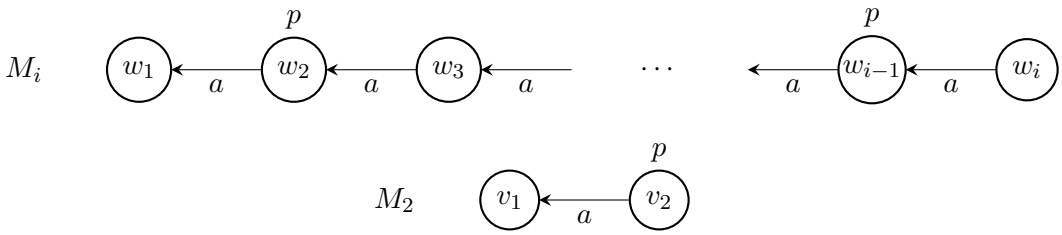

Figure 9: According to Demey's notion of bisimulation, model M_i (above) with alternating $\neg p$ and p worlds is a bisimulation contraction. In this particular case i is odd as p does not hold at w_i. According to our notion of bisimulation, all p worlds in model M_i are bisimilar and also all $\neg p$ worlds. Model M_2 (below) is the contraction.

Bisimulation Prior to our work Demey discussed the model theory of plausibility models in great detail in [13]. Our results add to the valuable original results he obtained. Demey does not consider degrees of belief; he considers knowledge, conditional belief and safe belief. Our plausibility models are what [13] refers to as uniform and locally connected epistemic plausibility models; he also considers models with fewer restrictions on the plausibility function. But given [13, Theorem 35], these types of models are for all intents and purposes equivalent to ours. The semantics for conditional belief and knowledge are as ours, but his semantics for safe belief is different (namely as in [7]). The difference is that in his case an agent safely believes φ if φ is true in all worlds as least as plausible as the current world, whereas in our case it is like that but *in the normalised model*. This choice of semantics has several highly significant implications as we will return to shortly.

In line with his interpretation of safe belief as a standard modality, Demey's notion of bisimulation for plausibility models is also standard. For example, whereas we require that

[forth$_\geq$] If $v \in W$ and $w \geq_a^R v$, $\exists v' \in W$ such that $w' \geq_a^R v'$ and $(v, v') \in R$,

where we recall that $w \geq_a^R v$ means $Min_a([w]_R \cap [w]_a) \geq_a Min_a([v]_R \cap [v]_a)$, he requires that

[forth$_\geq$] If $v \in W$ and $w \geq_a v$, $\exists v' \in W$ such that $w' \geq_a v'$ and $(v, v') \in R$.

He obtains correspondence for bisimulation and modal equivalence in the logic of safe belief in [13, Footnote 12 and Theorem 32]. Our notion of bisimulation is less restrictive, as we will now illustrate by way of the examples in Figure 9.

Consider model M_i in Figure 9. This is a single-agent model on a single proposition p containing i worlds, where the image of a world w_j under $\geq_a$ is $\{w_1, \ldots, w_j\}$. The valuation is such that if the index of a world is even then p holds, and otherwise p does not hold. Now, using Demey's notion of bisimulation entails that the largest autobisimulation on M_i is the identity, and thus M_i is a bisimulation contraction. For example, we can find a formula that distinguishes (M_i, w_i) from (M_{i+2}, w_{i+2}). For safe belief $\Box$ we now have Demey's semantics (see above) $M, w \models \Box_a \varphi$ iff $M, v \models \varphi$ for all v with $w \geq_a v$. We now define $\varphi_0 = \top$ and for any natural number $n \geq 1$ we let:

$$\varphi_n = \begin{cases} \Diamond_a(\varphi_{n-1} \wedge p) & \text{if } n \text{ is even;} \\ \Diamond_a(\varphi_{n-1} \wedge \neg p) & \text{if } n \text{ is odd;} \end{cases}$$

for example

$$\varphi_4 = \Diamond_a(\Diamond_a(\Diamond_a(\Diamond_a(\top \wedge \neg p) \wedge p) \wedge \neg p) \wedge p).$$

We now have that for any $i \geq 1$, $M_i, w_i \models \varphi_i \wedge \neg\varphi_{i+1}$, which makes this a distinguishing formula between (M_i, w_i) from (M_{i+2}, w_{i+2}). In fact, the semantics of $\Box_a$ allow us to count the number of worlds in M_i. In this sense Demey's logic is immensely expressive.

Again referring to Figure 9, consider M_3, the model with a most plausible $\neg p$ world, a less plausible p world and an even less plausible $\neg p$ world. In the logic L^C of conditional belief w_1 and w_3 of M_3 are modally equivalent. Hence they also ought to be bisimilar. But in Demey's notion of bisimilarity they are not. Hence we have a mismatch between modal equivalence and bisimilarity, which is not supposed to happen: it is possible for two worlds to be modally equivalent but not bisimilar. Demey also was aware of this, of course. To remedy the problem one can either strengthen the notion of modal equivalence or weaken the notion of bisimilarity. Demey chose the former (namely by adding the safe belief modality to the conditional belief modality), we chose the latter. Thus we regain the correspondence between bisimilarity and modal equivalence. Baltag and Smets [7] achieve the same via a different route: they include in the language special propositional symbols, so-called S-propositions. The denotation of an S-proposition can be any subset of the domain. This therefore also makes the language much more expressive.

We believe that in particular for application purposes, weakening the notion of bisimulation, as we have done, is preferable over strengthening the logic, as in [7, 13]. This come at the price of a more complex bisimulation definition (and, although we did not investigate this, surely a higher complexity of determining whether two worlds are bisimilar), but, we venture to observe, also a very elegant bisimulation definition given the ingenious use of the bisimulation relation itself in the definition of the forth and back conditions of bisimulation. We consider this one of the highlights of our work.

Expressivity In [7] one finds many original expressivity results. Our results copy those, but also go beyond. We recall Table 1 for the full picture of our results, and the main results of those namely $L^C < L^S$, $L^C \bowtie L^D$, and $L^D \bowtie L^S$. The first, $L^C < L^S$, is originally found in [7, page 34, Equation 1.7], and we obtained it using the same embedding translation. However, it may be worth to point out that in our case this translation still holds for the (in our opinion) more proper bisimulation preserving notion of safe belief. Baltag and Smets' S-propositions are arbitrary subsets of the domain, the (unnecessarily) far more expressive notion of safe belief. Baltag and Smets also discuss degrees of belief but do not obtain expressivity results for that, so $L^C \bowtie L^D$ may be considered novel and interesting. In artificial intelligence, the degrees of belief notion seems more widely in use than the conditional belief notion, so an informed reader had better be aware of the incomparability of both logics and may choose the logic to suit his or her needs. The result that $L^D \bowtie L^S$ could possibly also be considered unexpected, and therefore valuable.

Planning An application area of plausibility models is epistemic planning. A consequence of Demey's notion of bisimulation is that even for single-agent models on a finite set of propositions, the set of distinct, contraction-minimal pointed plausibility models is infinite. For example, we recall that in Figure 9 any two pointed plausibility models in $\{(M_i, w_i) \mid i \in \mathbb{N}\}$ are non-bisimilar. With our notion of bisimulation, there are in the single-agent case only finitely many distinct pointed plausibility models up to bisimulation. This was already reported in [4]. Our motivation for this bisimulation investigation was indeed prompted by the application of doxastic logics in planning.

In planning, an agent attempt to find a sequence of action, a plan, that achieves a given goal. A planning problem implicitly represents a state-transition system, where transitions are induced by actions. By exploring this state-space we can reason about actions and synthesise plans. A growing community investigates planning by applying dynamic epistemic logics [10, 26, 1], where actions are epistemic actions. Planning with doxastic modalities has also been considered [2]. This is done by identifying states with (pointed) plausibility models, and the goal with a formula of the doxastic language. Epistemic actions can be public actions, like hard and soft announcements [32], but also non-public actions, such as event models [7].

With the state-space consisting of plausibility models, model theoretic results become pivotal when deciding the plan existence problem. Unlike Demey's approach, our framework leads to a finite state-space in the single-agent case and therefore the single-agent plan existence problem is decidable [10]. At the same time we know that even in a purely epistemic setting the multi-agent plan existence problem is undecidable [10]. But by placing certain restrictions on the planning problem it is possible to find decidable fragments even in the multi-agent case, for example, event models with propositional preconditions [41].

Acknowledgements

We thank Giovanni Cina and Johannes Marti for productive exchanges of ideas following an ILLC seminar in 2015. We are also in dept to the anonymous reviewers for their thorough reading of the manuscript leading to many helpful comments and suggestions for revisions. Hans van Ditmarsch is also affiliated to IMSc (Institute of Mathematical Sciences), Chennai, as research associate. He acknowledges support from European Research Council grant EPS 313360. Preliminary versions of the results in this paper can be found in the PhD theses of Mikkel Birkegaard Andersen [3, Chapter 4] and Martin Holm Jensen [18, Chapter 5].

References

[1] M.B. Andersen, T. Bolander, and M.H. Jensen. Conditional epistemic planning. In *Proc. of 13th JELIA*, LNCS 7519, pages 94–106. Springer, 2012.

[2] M.B. Andersen, T. Bolander, and M.H. Jensen. Don't plan for the unexpected: Planning based on plausibility models. *Logique et Analyse*, 2015, To Appear.

[3] Mikkel Birkegaard Andersen. *Towards Theory-of-Mind agents using Automated Planning and Dynamic Epistemic Logic*. PhD thesis, Technical University of Denmark, 2015.

[4] Mikkel Birkegaard Andersen, Thomas Bolander, Hans P. van Ditmarsch, and Martin Holm Jensen. Bisimulation for single-agent plausibility models. In Stephen Cranefield and Abhaya Nayak, editors, *Australasian Conference on Artificial Intelligence*, volume 8272 of *Lecture Notes in Computer Science*, pages 277–288. Springer, 2013.

[5] G. Aucher. A combined system for update logic and belief revision. In *Proc. of 7th PRIMA*, pages 1–17. Springer, 2005. LNAI 3371.

[6] A. Baltag and S. Smets. The logic of conditional doxastic actions. In *New Perspectives on Games and Interaction*, Texts in Logic and Games 4, pages 9–31. Amsterdam University Press, 2008.

[7] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In *Proc. of 7th LOFT*, Texts in Logic and Games 3, pages 13–60. Amsterdam University Press, 2008.

[8] Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. In Giacomo Bonanno, Wiebe van der Hoek, and Michael Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT7)*, volume 3 of *Texts in Logic and Games*, pages 13–60. Amsterdam University Press, 2008.

[9] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge, UK, 2001.

[10] T. Bolander and M.B. Andersen. Epistemic planning for single and multi-agent systems. *Journal of Applied Non-classical Logics*, 21(1):9–34, 2011.

[11] Craig Boutilier. Conditional logics of normality as modal systems. In *AAAI*, volume 90, pages 594–599, 1990.

[12] K. Britz and I. Varzinczak. Defeasible modalities. In *Proc. of the 14th TARK*, 2013.

[13] L. Demey. Some remarks on the model theory of epistemic plausibility models. *Journal of Applied Non-Classical Logics*, 21(3-4):375–395, 2011.

[14] Kit Fine. In so many possible worlds. *Notre Dame Journal of Formal Logic*, 13(4):516–520, 1972.

[15] N. Friedman and J.Y. Halpern. A knowledge-based framework for belief change - part i: Foundations. In *Proc. of 5th TARK*, pages 44–64. Morgan Kaufmann, 1994.

[16] A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[17] J.Y. Halpern. *Reasoning about Uncertainty*. MIT Press, Cambridge MA, 2003.

[18] Martin Holm Jensen. *Epistemic and Doxastic Planning*. PhD thesis, Technical University of Denmark, 2014.

[19] S. Kraus and D. Lehmann. Knowledge, belief and time. *Theoretical Computer Science*, 58:155–174, 1988.

[20] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[21] Philippe Lamarre. S4 as the conditional logic of nonmonotonicity. *KR*, 91:357–367, 1991.

[22] N. Laverny. *Révision, mises à jour et planification en logique doxastique graduelle*. PhD thesis, Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France, 2006.

[23] W. Lenzen. Recent work in epistemic logic. *Acta Philosophica Fennica*, 30:1–219, 1978.

[24] W. Lenzen. Knowledge, belief, and subjective probability: outlines of a unified system of epistemic/doxastic logic. In V.F. Hendricks, K.F. Jorgensen, and S.A. Pedersen, editors, *Knowledge Contributors*, pages 17–31, Dordrecht, 2003. Kluwer Academic Publishers. Synthese Library Volume 322.

[25] D.K. Lewis. *Counterfactuals*. Harvard University Press, Cambridge (MA), 1973.

[26] B. Löwe, E. Pacuit, and A. Witzel. DEL planning and some tractable cases. In *Proc. of LORI 3*, pages 179–192. Springer, 2011.

[27] T.A. Meyer, W.A. Labuschagne, and J. Heidema. Refined epistemic entrenchment. *Journal of Logic, Language, and Information*, 9:237–259, 2000.

[28] K. Segerberg. Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic*, 39(3):287–306, 1998.

[29] W. Spohn. Ordinal conditional functions: a dynamic theory of epistemic states. In W.L. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change, and Statistics*, volume II, pages 105–134, 1988.

[30] R. Stalnaker. Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy*, 12:133–163, 1996.

[31] J. van Benthem. One is a lonely number: on the logic of communication. In *Logic colloquium 2002. Lecture Notes in Logic, Vol. 27*, pages 96–129. A.K. Peters, 2006.

[32] J. van Benthem. Dynamic logic of belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007.

[33] J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2011.

[34] W. van der Hoek. On the semantics of graded modalities. *Journal of Applied Non-Classical Logics*, 2(1), 1992.

[35] W. van der Hoek. Systems for knowledge and beliefs. *Journal of Logic and Computation*, 3(2):173–195, 1993.

[36] H. van Ditmarsch. Prolegomena to dynamic logic for belief revision. *Synthese (Knowledge, Rationality & Action)*, 147:229–275, 2005.

[37] H. van Ditmarsch. Comments on 'The logic of conditional doxastic actions'. In *New Perspectives on Games and Interaction*, Texts in Logic and Games 4, pages 33–44. Amsterdam University Press, 2008.

[38] H. van Ditmarsch, D. Fernández-Duque, and W. van der Hoek. On the definability of simulation and bisimulation in epistemic logic. *Journal of Logic and Computation*, 2012. doi:10.1093/logcom/exs058.

[39] H. van Ditmarsch and W.A. Labuschagne. My beliefs about your beliefs – a case study in theory of mind and epistemic logic. *Synthese*, 155:191–209, 2007.

[40] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*, volume 337 of *Synthese Library*. Springer, 2007.

[41] Quan Yu, Ximing Wen, and Yongmei Liu. Multi-agent epistemic explanatory diagnosis via reasoning about actions. In Francesca Rossi, editor, *IJCAI*. IJCAI/AAAI, 2013.